INTRODUCTION TO THE
GLOBAL STRUCTURE OF SPACE-TIME

THEORETICAL FOUNDATIONS OF COSMOLOGY

MICHAEL HELLER

World Scientific

INTRODUCTION TO THE
GLOBAL STRUCTURE OF SPACE-TIME

THEORETICAL FOUNDATIONS OF COSMOLOGY

Digitized by the Internet Archive
in 2022 with funding from
Kahle/Austin Foundation

https://archive.org/details/theoreticalfound0000hell

INTRODUCTION TO THE
GLOBAL STRUCTURE OF SPACE-TIME

THEORETICAL FOUNDATIONS OF COSMOLOGY

MICHAEL HELLER

Pontifical Academy of Theology
Cracow, Poland
and
Vatican Observatory, Vatican City

World Scientific
Singapore • New Jersey • London • Hong Kong

Published by

World Scientific Publishing Co. Pte. Ltd.
P. O. Box 128, Farrer Road, Singapore 9128
U.S.A. office: Suite 1B, 1060 Main Street, River Edge, NJ 07661
U.K. office: 73 Lynton Mead, Totteridge, London N20 8DH

THEORETICAL FOUNDATIONS OF COSMOLOGY
Introduction to the Global Structure of Space-Time

Copyright © 1992 by World Scientific Publishing Co. Pte. Ltd.

All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the Publisher.

ISBN 981-02-0756-5

Typeset by Z. Golda,
supported by the Polish Grant GMEN 141.

Printed in Singapore by JBW Printers & Binders Pte. Ltd.

Table of Contents

Preface vii

Chapter 1. Manifold Model of Space–Time 1

0. Introduction 1

1. Banach Spaces and Manifolds — Mathematical Environment of the Space–Time Theory 2

2. Differential Manifolds 5

3. Orientability of Manifolds 10

4. Vector and Tensor Fields on Manifolds 11

5. Submanifolds 14

6. Manifold as a Model of Space–Time 15

7. Comments and Remarks 16

Bibliographical Note 19

Chapter 2. Relativistic Model of Space–Time 21

0. Introduction 21

1. Metric Manifolds 22

2. Metric Model of Space–Time 24

3. Space–Time of Special Relativity 25

4. The Existence of the Lorentz Structure 27

5. Comments and Remarks 30

Bibliographical Note 32

Chapter 3. Method of Fibre Bundles 34

0. Introduction 34

1. Fibre Bundles 36

2. Principal and Associated Fibre Bundles 39

3. Connections in a Principal Fibre Bundle 41

4. Geometry of Principal Fibre Bundles 45

5. Geometry of the Frame Bundle 47

6. Applications to Gauge Theories 49

Bibliographical Note 51

Chapter 4. The Frame Bundle and General Relativity 52

0. Introduction 52

1. Geometry of the Frame Bundle (continuation) 53

2. Geometry of the Base Manifold 56
3. Geometry of the Base Manifold (continuation) 58
4. The Fibre Bundle of Orthonormal Frames 59
5. General Relativity 61
6. Comments and Remarks 63
Bibliographical Note 65
Chapter 5. The Cosmological Problem 66
0. Introduction 66
1. Cosmology as a Global Geometry of Space–Time 67
2. Cosmology as a Theory of the Ensemble of Universes 70
3. Observational Problems of Cosmology 76
Chapter 6. Geometry of the Standard Cosmological Model 83
0. Introduction 83
1. Warped Products 84
2. Macroscopic Description of Particles and Observers 86
3. Construction of the Standard Cosmological Model 88
4. Evolution of the Standard Cosmological Model 90
5. Dust Filled and Radiation Filled Universes 93
6. Observational Background of the Standard Cosmological Model . . . 94
7. Comments and Remarks 101
Bibliographical Essay 106
Appendix. Evolution of Dynamical Space–Time Theories 112
0. Introduction 112
1. Elementary Repetition of the Category Theory 113
2. Aristotelian Space–Time 117
3. Space–Time of Classical Mechanics Without Gravitational Field 118
4. Space–Time of Classical Mechanics with Gravitational Field 121
5. The Special and General Theories of Relativity 122
6. Comments and Remarks 122
Bibliographical Note 125
References 126
Symbol Index 135
Author and Subject Index 141

Preface

The aim of the present book is to discuss theoretical foundations of cosmology. And nothing more, but possibly also nothing less. Therefore, it is not a textbook in any sense of this word. One of its principal goals is to study the assumptions upon which relativistic cosmology is founded, and, first of all, those assumptions that are implicit in the mathematical structure of cosmology.

The leading idea of the book may be formulated in the following way: *The Universe is a structure*. This structure can be investigated with the help of either local or non–local (global) mathematical methods. In the first case, we have theoretical physics, in the traditional meaning of this word (i.e. in the sense of "local physics"). In the second case, we have cosmology. In this sense, *cosmology is a non–local physics*. In the spirit of this approach, I use, from the very beginning, global mathematical methods, even in these parts of the material which are usually regarded as only a preparation to cosmology.

It is worth noticing that cosmology, understood as a non–local physics, is something more than a theory of cosmological models. Cosmology in such a narrow sense (as a theory of world models) appears in the book not earlier than in the last two chapters; I think, however, that, when analysing the global structure of space–time in all foregoing chapters, I am studying truly cosmological issues. Both special and general theories of relativity are presented with stress on their respective global aspects.

I assume that the reader is familiar with mathematical methods which are employed throughout the book; although, for his or her convenience, I usually give main definitions with short comments and some examples. In this way, I hope, the reading of the book could be smooth enough without having too often to consult other textbooks or monographs. However, I am aware that more responsible acquaintance with the book would sometimes require the aid of some other publications. *Bibliographical notes*, at the end of every chapter (with the exception of chapters 5 and 6), could assist the reader in this respect.

I would like to stress once more that my intention is not to teach the reader how to do calculations (I assume that he already knows it), but to clarify concepts and show the efficiency of structures.

Definitions, theorems, examples, and so on, are numbered in a continuous manner within each section; the method of numbering is natural and needs no explanation. Only the definitions of those concepts which play a role in the exposition have been numbered, definitions of all other concepts having been incorporated into the text.

The organisation of the material is the following. In the first two chapters foundations and assumptions of the contemporary theory of space–time are discussed. The second chapter contains elements of special relativity. I am especially interested in the existence of the Lorentz structure. Since it is, in principle, a global problem, it conveys — in my opinion — information of cosmological significance. The third chapter is a concise repetition of the fibre bundle method, one of the main theoretical tools

used in the book. With the help of it the general theory of relativity is presented in the fourth chapter. The method, on the one hand, makes manifest similarities of this theory with gauge theories of contemporary physics and, on the other hand, stresses the peculiarity of Einstein's theory of gravity as a theory of the frame bundle over space–time. The fifth chapter occupies the central place in the book. It poses, to use Einstein's expression, "the cosmological problem", and explicitly formulates the leading idea of the entire book. Although no single formula appears in this chapter, it is by no means a popular exposition: its full understanding presupposes a considerable mathematical and cosmological culture by the reader. Analyses carried out in this chapter (and a program it outlines) are applied to the simplest world models, the so–called Robertson–Walker cosmology, in the sixth chapter. It goes without saying that the Robertson–Walker cosmology is reconstructed with the help of global methods (the concept of warped product plays a key role here).

The rather extensive Appendix (of the size of an average chapter) is aimed at revealing a beautiful logic inherent in the evolution of space–time theories, starting from the Aristotelian physics and ending with general relativity. The application of the fibre bundle method, helped — if necessary — by some concepts of the mathematical category theory, has allowed us to reconstruct former theories of space–time in the modern geometric framework, and to show a "natural" character of the Einstein theory of relativity within the context of this evolutionary chain. It turns out that general relativity is logically the almost unavoidable "next step" along the path of development. In my opinion, it is a strong, albeit purely esthetic, argument on behalf of Einstein's theory.

At the end of most of the chapters there is a section entitled *Comments and Remarks*. Here I consider some problems more loosely connected with the content of the given chapter, and sometimes I allow myself digressions of a philosophical character.

I am fully aware that the book leaves many problems either completely untouched or non–satisfactorily treated. Such problems as: Bianchi cosmology, physics of the early Universe, further chapters of the global theory of space–time (including the classical singularity problem), many questions of observational cosmology, although well (and sometimes very well) developed as far as their technical aspects are concerned, still await a more careful methodological elaboration. Each of these problems (and many others as well) deserve a separate monograph. The *Bibliographical Essay*, after the sixth chapter, containing a list of books on cosmological topics with short remarks and explanations, is intended to recompense the reader for the fact that the book ends at a point where the most interesting action only begins.

M. H.

March 1985

INTRODUCTION TO THE
GLOBAL STRUCTURE OF SPACE-TIME

THEORETICAL FOUNDATIONS OF COSMOLOGY

Chapter 1

Manifold Model of Space–Time

0. Introduction

Each physical theory contains cosmological elements, usually as tacit assumptions, since each physical theory develops on an "arena", and the theory of such an arena can be thought of as a part of cosmology. Macroscopic physical theories occur in space-time and all of them presuppose a certain structure of it, called *differential manifold structure* or simply *manifold structure*. This structure is common to all macroscopic physical theories which differ from one another only when structures are taken into account that are superimposed on the manifold structure. The problem of a suitable arena for theories of microphysics is still an open question. Quite a common opinion asserts that, in quantum mechanics also, manifold continues to play its role of an arena for physical processes. Even if the opponents of this opinion are right, one has to admit that space-time modelled by differential manifold preserves its importance in all quantum theories: all measurements that constitute the empirical base of such theories are performed within the macroscopic space–time, and they must appear, in this or another way, in the mathematical formalism of these theories (here all discussions have their origin concerning the theoretical status of space–time in theories of microphysics).

To construct the manifold model of space time is the goal of the present chapter. Its larger parts are of a purely mathematical character. Sec. 1 prepares, so to speak, the mathematical environment for the theory of differential manifolds. It presents, in a more or less informal manner, some information on functional spaces, and in particular on Banach spaces, and their role in modern differential geometry as well as some preliminary remarks concerning the geometric model of space–time. Sec. 2 contains the definition of differential manifold (in the case of a finite number of dimensions) and introduces the most important concepts of the manifold theory. Sec. 3 deals with one of such concepts, namely the concept of manifold orientability. Vector and tensor fields are introduced in sec. 4, and the notions of immersion, embedding and submanifolds appear in sec. 5. Mathematical tools, prepared in this way, enable us to present in sec. 6 a succinct construction of the manifold model of space–time. Sec. 7 contains comments and remarks more loosely connected with the main topic of the chapter. Among others, the attention of the reader is drawn to some difficulties inherent in the very concept of space-time manifold, which made impossible the full implementation into a physical theory the so called Mach's principle, once postulated

by Einstein and broadly discussed till now. At the end of sec. 7 some remarks can be found concerning the status of space–time manifold in theories of the microworld.

Mathematical material presented in this chapter, as well as in a few next chapters, is of a repetitive character, therefore definitions are stressed rather than theorems and their proofs, mutual dependencies between concepts that will be needed later on rather than particular examples. In the "Bibliographical Note" at the end of the chapter, the reader can find information concerning books on the manifold theory. He is encouraged to choose something from that list. Clear mathematical perspective will make further reading of this book easier.

The model of space–time, as a four -dimensional differential manifold, presented in this chapter is very general. It will be gradually enriched in the following chapters.

1. Banach Spaces and Manifolds — Mathematical Environment of the Space–Time Theory

The general concept of function appeared in mathematics not earlier than the beginning of the XIX$^{\text{th}}$ century. It was Riemann who, in his famous inaugural lecture *On the Hypotheses That Lie at the Foundations of Geometry*, first noted the necessity to investigate functional spaces. Further development of the function theory consisted mainly in a transition to the study of functions whose domains and ranges are not limited to the axis of real numbers or to spaces with finite numbers of dimensions. Owing to this transition, functional spaces were put into their natural environment, namely the general theory of topological spaces.

Dwelling within the region of elementary mathematics creates a dangerous illusion. One usually deals with "well–behaved" functions, for instance, with continuous or differentiable (as many times as needed) functions, with functional spaces having sufficiently smooth structures, etc. This creates an impression that well–behaved functions exist in general, and "pathological cases", such as noncontinuous or nondifferentiable functions, are rare exceptions. An instant of reflection (for which seldom a time is left in a spate of computations) is enough to see that things are just the opposite. There are well–behaved functions that constitute exceptional regions (of "measure zero") within the realm of all possible functions.

Non–maliciousness of God, of which Einstein spoke so often, reveals itself in that in his architectural plans the Creator has used well–behaved functions or at least such functions that can be approximated by well–behaved functions. The functions being elements of the Banach space belong to this privileged class.

A linear normed and complete space is said to be a *Banach space*. This definition can be thought of as a natural generalization of the concept of vector field in the Euclidean space. To each element x of a Banach space (to each "vector") one ascribes a real number $\|x\|$, called the *norm* of the element (a generalization of a vector's length). The norm is defined with the help of known axioms that formalise its properties. Linearity of a Banach space ensures that one can add its elements

to each other and multiply them by numbers (real or complex). Completeness of a Banach space means that every Cauchy sequence reaches its limit in this space.

The properties expressed in the above definition of Banach space guarantee that, on the one hand, this space is a very general object containing a rich class of special cases often met in the mathematical practice and, on the other hand, it is a well behaved object eliminating many pathological situations which would require special caution from the mathematician.

In particular, from the definition of Banach space it follows that a metric can be introduced in this space. Indeed, the expression $\|x_1 - x_2\|$, where x_1 and x_2 are elements of a Banach space, satisfies the metric axioms. Therefore, Banach spaces are metrizable.

One of the main advantages of Banach spaces is that they constitute a natural environment for calculus. The generalization of the derivative in the space $\mathbf{R}^n$ (which is a special case of a Banach space) to the derivative in Banach spaces in general is so immediate that many theorems, obtained in this way, look as if they were mechanically repeated from the elementary course of differential calculus. In fact, this procedure is far from being trivial. In the transition process to general Banach spaces, the notion of a derivative of a function is generalized to the concept of a mapping from one Banach space to another. In this way, functional analysis acquires one of its most fundamental tools.

Calculus can also be done in spaces more general than Banach spaces, for instance, in any topological vector space. However, in such a case, the generalization is not that natural (for example, the implicit function theorem can no longer be valid without additional assumptions).

The theory of functional spaces is a powerful tool of modern differential geometry. Differential geometry has grown up from analytic investigations of curves and surfaces in the three–dimensional Euclidean space. The generalization went in two directions: first, from three to any number of dimensions (also infinite) and, second, towards getting rid of the "surrounding" Euclidean space. In this way, the crucial concept of modern differential geometry was born, namely the concept of *differential manifold modelled on a topological vector space of any (also infinite) number of dimensions* or, in particular, *on a Banach space*. In the latter case one speaks about a *Banach manifold.*

The manifold concept is a generalization of the parametric representation of a surface in $\mathbf{R}^3$, i.e. of a $\mathcal{C}^k$ mapping of an open subset of a given subset into $\mathbf{R}^2$. In the case of an infinite-dimensional Banach manifold, the space $\mathbf{R}^2$ is replaced by an infinite–dimensional Banach space.

Let X be any set and $\mathcal{B}$ a Banach space. The set of pairs (U_i, ϕ_i), called *local charts*, such that

(1) the sets U_i $(i \in \mathcal{I})$ cover X,

(2) the mappings ϕ_i are $\mathcal{C}^k$ bijections from open subsets U_i onto open subsets of $\mathcal{B}$,

(3) the sets $\phi_i(U_i \cap U_j)$, for each pair $i, j \in \mathcal{I}$, are open in $\mathcal{B}$, and mappings of the

type

$$\phi_j \circ \phi_i^{-1}\colon\ \phi_i(U_i \cap U_j) \longrightarrow \phi_j(U_i \cap U_j)$$

are of $\mathcal{C}^k$ class,

is called *the $\mathcal{C}^k$ atlas on X modelled on the space $\mathcal{B}$*. Two atlases are called *equivalent* if their union is an atlas. The equivalence of atlases is an equivalence relation.

A set X together with an equivalence relation of $\mathcal{C}^k$ atlases on X modelled on the Banach space $\mathcal{B}$ is said to be a *$\mathcal{C}^k$ differential manifold modelled on the Banach space $\mathcal{B}$*. The dimension of such a differential manifold is, from definition, the dimension of $\mathcal{B}$; it can be infinite. The atlases that belong to the equivalence class appearing in the above manifold definition are called atlases *admissible on X*.

Infinite dimensional manifolds modelled on Banach spaces, in short also called *Banach manifolds*, are in fact functional spaces; however, one can introduce on them a majority of structures which are natural generalizations of structures known from the elementary course of differential geometry. For instance, tangent and cotangent spaces, and consequently vector and tensor fields as well as differential forms, can be defined on them. In particular, symplectic forms and Hamiltonian vector fields are naturally introduced on Banach manifolds. In close analogy to the finite–dimensional case one can construct various fibre bundles over Banach manifolds and define fields of different geometric objects as cross–sections of the corresponding fibre bundles over these manifolds. Tensor fields and differential forms turn out to be special cases of such constructions. In this way, one can also introduce a metric (the field of a metric tensor) on a Banach manifold, changing it into a *Riemann manifold*. The geometry of such manifolds is further constructed in analogy with the finite–dimensional case.

The spaces $\mathbf{R}^n$ (and $\mathbf{C}^n$) are Banach spaces with the norm defined in a natural way, and differential manifolds modelled on $\mathbf{R}^n$ (or $\mathbf{C}^n$), (their precise definitions will be given in the next section), play a key role in the traditional differential geometry; they are simply called *differential manifolds* (*real* or *complex*). Differential manifolds are locally diffeomorphic to $\mathbf{R}^n$ (or $\mathbf{C}^n$), and, therefore, at each of their points there exists tangent vector space with the tangent vectors defined in an "intrinsic manner", i.e. with no reference to any space in which the manifold would be immersed. On the one hand, differential manifolds preserve well–known properties of surfaces immersed in the Euclidean spaces, and, on the other hand, they are entirely autonomous in the sense that, although modelled on $\mathbf{R}^n$ (or $\mathbf{C}^n$), they do not need to be considered with reference to any "surrounding space".

Owing to these properties, in the contemporary theoretical physics four–dimensional (real) differential manifolds serve as models of space–time.[1] This remains true both with respect to classical physics and with respect to special and general rela-

[1]The reader should pay attention to a certain fuzziness of commonly accepted terminology: one speaks of manifolds **modelled on** Banach spaces, and one speaks on the physical space–time **modelled by** a four-dimensional manifold. Modelling in the former sense is determined by the manifold definition; modelling in the latter sense is one of the most fundamental procedures used in theoretical physics, whose methodological aspects are studied by philosophers of science.

tivity. It turns out that space–times of pre–Newtonian theories (those of Aristotle and Galileo) can also be modelled by four-dimensional differential manifolds (see, Appendix). The manifold model of the physical space–time is a subject matter of the present book. In this introductory section, I want only to place this space–time model in its natural mathematical environment, namely in the general theory of differential manifolds modelled on Banach spaces. Real differential manifolds constitute a very special class of Banach manifolds, but nevertheless they are general enough to provide description tools of a rich family of phenomena occurring in the real universe. It seems that a certain tension between generality and peculiarity of mathematical constructions is a key characteristic of the conceptual structure of contemporary physics.

The concept of normed spaces was introduced by Stefan Banach in 1922 in his doctoral thesis. From Banach's work, *Theorie des operations lineaires* (1932) the "mature age of normed spaces" began (Bourbaki 1969). The manifold concept started to crystalise itself in the XIX$^{\text{th}}$ century, simultaneously in several branches of mathematics, where the necessity to investigate locally Euclidean spaces arose. In 1854, in his outspoken inaugural lecture, Riemann made some remarks which turned out to be important for the development of the manifold concept (see Riemann 1868). The next step belongs to Poincaré: in 1880, when studying certain topological techniques in the theory of differential equations, he introduced what is to–day called flows on a manifold. A precise definition of a smooth two–dimensional manifold was given by Weyl in 1913, in his book *The Concept of a Riemann Surface* (see Weyl 1955). Understanding of the manifold concept evolved rather laboriously in the first half of the XX$^{\text{th}}$ century to become, after 1950, one of the most crucial chapters of modern topology and differential geometry.

2. *Differential Manifolds*

A topological space X that is homeomorphic to $\mathbf{R}^n$ in the neighbourhood of each of its points is said to be a *topological manifold*. In the following, we shall assume — unless this assumption is directly revoked — that topological manifolds have the *Haussdorf property*, i.e., any two different points of such a topological manifold possess disjoint neighbourhoods. The existence of a local homeomorphism between X and $\mathbf{R}^n$ should be understood in the sense of the following definitions.

A pair (U, ϕ), where U is an open subset of X, and ϕ: $U \to V$ a homeomorphism onto an open subset $V \subset \mathbf{R}^n$ is said to be a (*local*) *chart* on the manifold X. By definition n is the *dimension* of the topological manifold. Coordinates $(x^1, \ldots, x^n)$ of the image $\phi(x) \in \mathbf{R}^n$ of a point $x \in X$ are called (*local*) *coordinates* of the point x in the chart (U, ϕ); this chart is also traditionally called a (*local*) *coordinate system on* X. It is assumed that X is covered by open sets U_α being domains of charts on X.

In the following, we shall additionally assume that the set $\{U_\alpha\}$ covering X is countable (X as a topological space has a countable base). However, this assumption is not a part of the manifold definition. Because of future applications it is convenient to accept that X is a connected space, or at least that the dimension of X is the same in the neighbourhood of each of its points (this property follows from the connectedness of X).

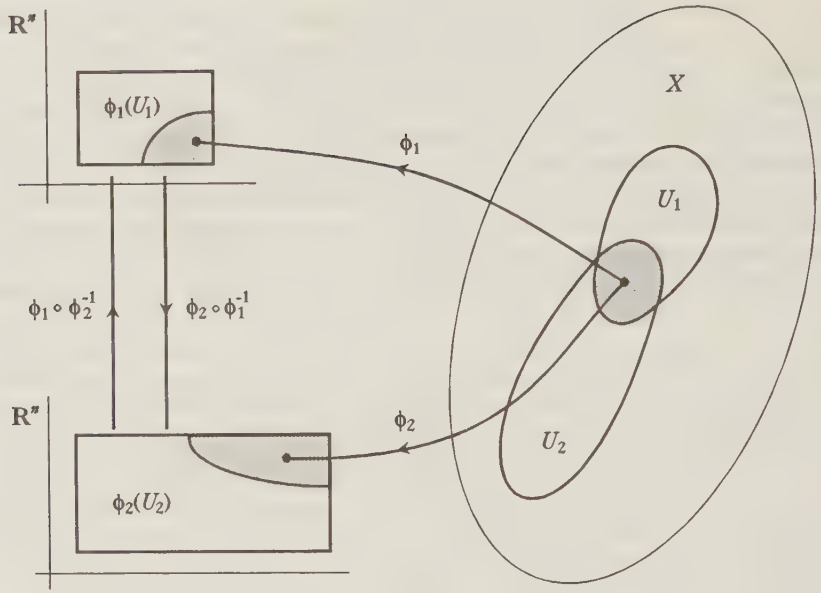


Fig. 1.1. Compatibility of maps on topological manifold.

Two charts (U_α, ϕ_α) and (U_β, ϕ_β) on X are said to be $\mathcal{C}^k$ *consistent* if — assuming that $U_\alpha \cap U_\beta \neq \emptyset$ — the mappings

$$\phi_\beta \circ \phi_\alpha^{-1}\colon\ \phi_\alpha(U_\alpha \cap U_\beta) \longrightarrow \phi_\beta(U_\alpha \cap U_\beta),$$
$$\phi_\alpha \circ \phi_\beta^{-1}\colon\ \phi_\beta(U_\alpha \cap U_\beta) \longrightarrow \phi_\alpha(U_\alpha \cap U_\beta)$$

from open subsets of $\mathbf{R}^n$ into open subsets of $\mathbf{R}^n$ are of class $\mathcal{C}^k$ (see Fig. 1.1). The compatibility condition means that there exists a smooth (in the sense of $\mathcal{C}^k$) transition from coordinates (x^i), $i = 1, \ldots, n$, defined on U_α, to coordinates (y^i) defined on U_β, in the region on which U_α and U_β overlap, i.e.,

$$(x^i) \longmapsto (y^i) = f^i(x^i),$$

where f^i are real functions of class $\mathcal{C}^k$ (and analogously the other way round: $(y^i) \mapsto (x^i)$). k may be also infinite. If the mappings $\phi_\beta \circ \phi_\alpha^{-1}$ $(\phi_\alpha \circ \phi_\beta^{-1})$ are analytic, we write $k = \omega$.

The set of charts $\{(U_\alpha, \phi_\alpha)\}$ covering X (i.e. $X = \cup_\alpha U_\alpha$) and pairwaise $\mathcal{C}^k$ consistent on overlapping domains, is said to a $\mathcal{C}^k$ *atlas A on X*. Two $\mathcal{C}^k$ atlases are said to be $\mathcal{C}^k$ *equivalent* if their union is a $\mathcal{C}^k$ atlas on X.

2.1. Definition. The pair $\mathcal{M} = (X, \{A\})$, where X is a topological space, and $\{A\}$ a set (an equivalence class, in fact) of equivalent $\mathcal{C}^k$ atlases on X, is said

to be a $\mathcal{C}^k$ *differential manifold* (or $\mathcal{C}^k$ manifold, for short). If there is no danger of misunderstanding, we will often omit the indication of the differentiability class ($\mathcal{C}^k$) in the name of manifolds. If $k = \omega$, a manifold is said to be *analytical*.

In practice, one works with one atlas treating it as a representation of the entire equivalence class. We shall often make use of this simplification.[2]

2.2. Example. The space $X = \mathbf{R}^n$ can be given differential structure by defining an atlas consisting of the one chart (U_α, ϕ), where $U = \mathbf{R}^n$, $\phi = \mathrm{id}$.

2.3. Example. To define a differential structure on a one-dimensional sphere

$$\mathcal{S}^1 = \{(x^1, x^2) \in \mathbf{R}^2 \colon (x^1)^2 + (x^2)^2 = 1\},$$

an atlas is needed consisting of at least of two charts (U_α, ϕ_α), $\alpha = 1, 2$, (see Fig. 1.2)

$$\begin{aligned} U_1 &= \mathcal{S}^1 \setminus \{(-1,0)\}, \qquad & \phi_1 &\colon U_1 \longrightarrow \mathcal{D}_1 \subset \mathbf{R}^1, \\ U_2 &= \mathcal{S}^1 \setminus \{(1,0)\}, \qquad & \phi_2 &\colon U_2 \longrightarrow \mathcal{D}_2 \subset \mathbf{R}^1. \end{aligned}$$

The mappings ϕ_α can be defined in the following way,

$$\begin{aligned} \phi_1^{-1} &\colon \Theta \longmapsto (\cos\Theta, \sin\Theta), \qquad & -\pi < \Theta < \pi, \\ \phi_2^{-1} &\colon \Theta \longmapsto (\cos\Theta, \sin\Theta), \qquad & 0 < \Theta < 2\pi \end{aligned}$$

(see, Thirring 1978, pp. 10–11).

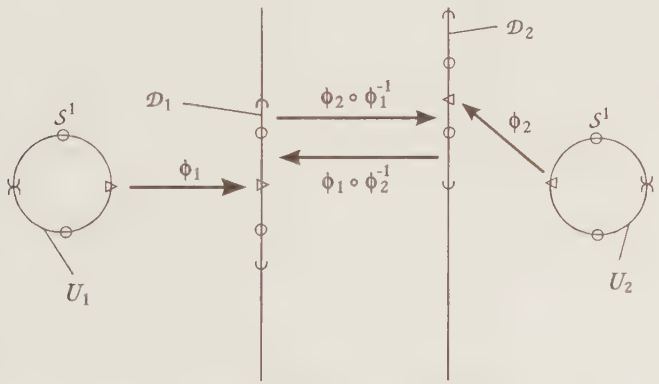


Fig. 1.2. Atlas on $\mathcal{S}^1$. For clarity the set $\mathcal{S}^1$ is shown in two copies. Little circles, triangles, and parentheses indicate which points correspond to each other.

[2]Against a more common usage we prefer the term *differential* rather than *differentiable* in the name $\mathcal{C}^k$ manifolds, understanding the difference between these two terms more or less in the same manner as one understands the difference between *metric* and *metrizable*. In this sense, not every topological manifold is differentiable, but if it is $\mathcal{C}^1$ differentiable, it is also $\mathcal{C}^\infty$ differentiable.

If, in the chain of definitions leading to the $\mathcal{C}^k$ manifold concept, the space $\mathbf{R}^n$ is replaced by its "upper" ("lower") part, i.e., by the region in $\mathbf{R}^n$ satisfying the condition $u^1 \geq 0$ ($u^1 \leq 0$), where u^1 is a coordinate in $\mathbf{R}^n$ (this region will be denoted by $\frac{1}{2}\mathbf{R}^n$), one obtains the so-called $\mathcal{C}^k$ *manifold with boundary*. The *boundary* $\partial\mathcal{M}$ (sometimes denoted also by $\dot{\mathcal{M}}$) of a manifold $\mathcal{M}$ with boundary is defined to be the set of those points of $\mathcal{M}$, the images of which under ϕ_α are situated on the boundary of $\frac{1}{2}\mathbf{R}^n$ in $\mathbf{R}^n$; the mappings ϕ_α define local coordinate systems on $\mathcal{M}$.

2.4. Example. Let $X = [a, b] \subset \mathbf{R}^1$. Any atlas on X must consist of at least two charts (U_α, ϕ_α), $\alpha = 1, 2$; for instance,

$$U_1 = [a, b), \qquad \phi_1 \colon x \longmapsto x - a,$$
$$U_2 = (a, b], \qquad \phi_2 \colon x \longmapsto b - x.$$

Then $\partial\mathcal{M} = \{a\} \cup \{b\}$.

Any $\mathcal{C}^k$ manifold can be also constructed in a different way, more clearly showing the role of the modelling space $\mathbf{R}^n$. The construction recipe is the following (see Bishop and Goldberg 1980, pp. 31–33). Prepare a suitable number of disjoint copies of $\mathbf{R}^n$, and from them cut out the open regions that will be modelling local coordinate systems on the manifold. The open regions which are to be overlapping with each other should be cut out of different copies of $\mathbf{R}^n$. Now identify those points of the open regions which are transformed into each other when the transition is made from one local coordinate system to another local coordinate system. To be more precise, on the set $\mathcal{P}$, being the union of disjoint open regions of different copies of $\mathbf{R}^n$, one defines an equivalence relation $\mathcal{E}$, treating as equivalent those points of different open regions which correspond to each other under local coordinate transformations. The manifold is to be defined as the quotient space $\mathcal{M} = \mathcal{P}/\mathcal{E}$. A topology on $\mathcal{M}$ can be introduced by demanding that local coordinate transformations be homeomorphisms. Such a topology is called a *manifold topology*. In short, a manifold can be glued together out of a sufficient number of pieces of $\mathbf{R}^n$. Whitney (1936) has proved that the two ways presented above of defining manifolds are equivalent.

2.5. Definition. Let $\mathcal{M} = (X^n, \{A\})$ and $\mathcal{N} = (Y^p, \{B\})$ be two $\mathcal{C}^k$ manifolds, and (U, ϕ) and (W, ψ) their local charts at points $x \in U$ and $y \in W$, respectively. Let us consider a mapping $f \colon X^n \to Y^n$ (Fig. 1.3). The expression $\psi \circ f \circ \phi^{-1}$ *represents* the mapping f in the charts (U, ϕ) and (W, ψ). The mapping f is $\mathcal{C}^r$ ($r \leq k$) at x if $\psi \circ f \circ \phi^{-1}$ is $\mathcal{C}^r$ at $\phi(x)$. It can be shown that this definition is independent of the choice of a chart. If a mapping f is $\mathcal{C}^r$ at each point $x \in X^n$, f is said to be of *class* $\mathcal{C}^r$ (or simply to be $\mathcal{C}^r$).

If $Y^p = \mathbf{R}^1$ (Fig. 1.4), f is called a *function* on X^n, and the expression $f \circ \phi^{-1}$ its *representative* in the chart (U, ϕ).

2.6. Definition. If a mapping $f \colon X^n \to Y^p$ (of class $\mathcal{C}^k$) is a bijection, and if both f and its inverse f^{-1} are continuously differentiable (are $\mathcal{C}^k$), f is called ($\mathcal{C}^k$) *diffeomorphism*. Two manifolds (of class $\mathcal{C}^k$) between which there *exists* a ($\mathcal{C}^k$) diffeomorphism are called ($\mathcal{C}^k$) *diffeomorphic*.

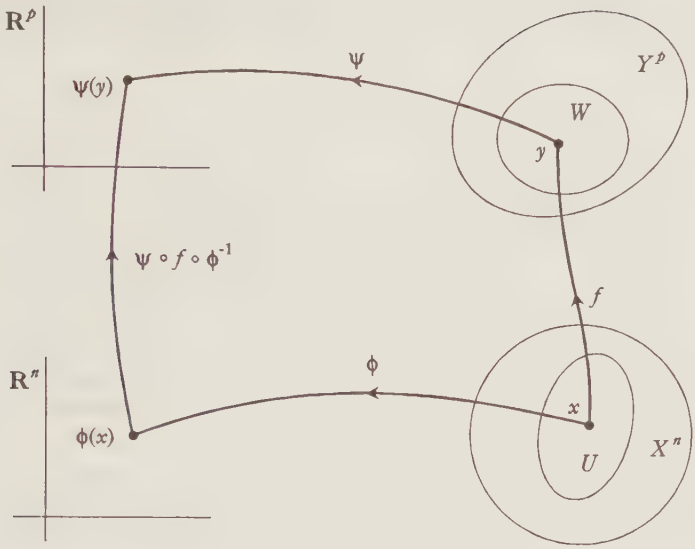


Fig. 1.3. Mapping $f: X^n \to Y^p$.

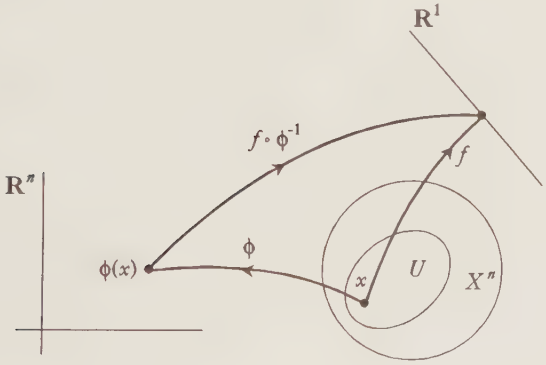


Fig. 1.4. Function f on manifold.

In the theory of manifolds, diffeomorphic manifolds are indistinguishable from each other; in other words, there exists a category, called the *category of $\mathcal{C}^k$ manifolds*, the objects of which are $\mathcal{C}^k$ manifolds and morphisms are differentiable mappings of class $\mathcal{C}^k$ between $\mathcal{C}^k$ manifolds. $\mathcal{C}^k$ diffeomorphisms are isomorphisms in this category. (See Trautman 1970b, 1973b; Sulanke and Wintgen 1972; concise information on the category theory is given in the Appendix, sec. 1.)

From the theorem proved by Whitney (1936) it follows that the theory of n–dimensional manifolds can be identified, in a sense, with the theory of subsets $\mathbf{R}^m$ with $m \leq 2n + 1$.

3. Orientability of Manifolds

Let (x^i) and (y^i) be local coordinate systems on open subsets of $\mathbf{R}^n$. The coordinate systems are said to have *the same orientation* if the determinant of the Jacobian matrix $\mathcal{J} = \det(\partial x^i / \partial y^j)$ is positive at each point of a given open subset of $\mathbf{R}^n$. The *orientation* of a chart (U, ϕ) of a manifold $\mathcal{M}$ is defined through the orientation of coordinate systems on $\phi(U) \subset \mathbf{R}^n$. A manifold is said to be *orientable* if there is an atlas on it such that any two charts (U_1, ϕ_1) and (U_2, ϕ_2), the domains of which overlap $(U_1 \cap U_2 \neq \emptyset)$, have the same orientation. A manifold defined in terms of such an atlas is said to be *oriented*. A manifold which is not orientable is said to be *nonorientable*.

Two atlases of a given manifold are said to possess *the same orientation* if all their charts have pairwaise the same orientations. To be of the same orientation is for the atlases an equivalence relation. An equivalence class of this relation is called an *orientation* of this atlas. Every connected manifold admits exactly two orientations.

The space $\mathbf{R}^n$, considered as a manifold with an atlas consisting of one chart $(\mathbf{R}^n, \mathrm{id})$ (see example 2.2), is an orientable manifold. The Möbius strip can serve as an example of a nonorientable manifold. In the case of two–sided surfaces in the three–dimensional Euclidean space, orientability is equivalent to two–sidedness of a surface (the Möbius strip is a one–sided surface). (For more information see Brickell and Clark 1970, pp. 119–125; Bishop and Goldberg 1980, pp. 162–164).

There exists a visual method of deciding whether a given manifold is orientable or not. Let us associate a *graph* $\mathcal{G}(A)$ with each atlas $A = \{(U_\alpha, \phi_\alpha)\}$ of a manifold $\mathcal{M} = (X, \{A\})$ in the following way. Let the knots of the graph $\mathcal{G}$ correspond to the charts ϕ_α of the atlas A, and the edges of the graph, linking the corresponding knots, to the connected components of the overlapping regions $U_\alpha \cap U_\beta \neq 0$. All these correspondencies are assumed to be one–to–one. Two knots of the graph can be linked by countably many edges. To all edges of the graph let us ascribe the numbers: $+1$ if the Jacobian matrix in the connected component of the region $U_\alpha \cap U_\beta$ that corresponds to that region is positive, and -1 if this Jacobian matrix is negative. In this way, each knot is linked with itself, forming a loop to which a number $+1$ is ascribed. The sequence of edges forming a loop will be called a *cycle*. To the sequences of edges (and consequently, to cycles as well) we ascribe a number, being the product of numbers which correspond to the edges forming this sequence (this cycle). We say that the *change of orientation of a knot* takes place if all numbers, corresponding to edges emanating from this knot, are multiplied by -1. Let us notice that this procedure does not change the number corresponding to the loop emanating from, and returning to, a given knot (this number is twice multiplied by -1). A graph

is said to be *orientable* if, by suitably changing the orientation of its knots, one can obtain the situation in which to all edges of the graph $\mathcal{G}$ the numbers $+1$ are ascribed.

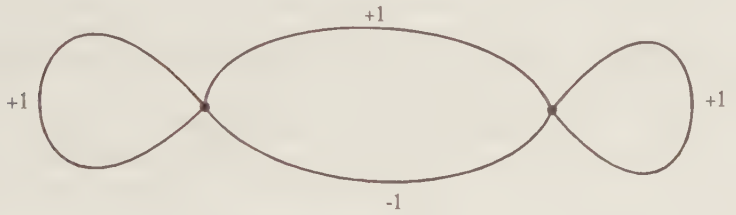


Fig. 1.5. Graph corresponding to the Möbius strip.

Of course, the change of orientation of a knot belonging to a graph $\mathcal{G}$ corresponds to the change of orientation of the chart ϕ_α associated with this knot and belonging to the atlas A of a given manifold $\mathcal{M}$. A manifold $\mathcal{M}$ is orientable if and only if the graph $\mathcal{G}$ corresponding to its atlas A is orientable. It can be shown that a graph $\mathcal{G}$ is oriented, and consequently the manifold corresponding to it is oriented, if $\mathcal{G}$ contains no cycle to which the number -1 is ascribed. (See Sulanke and Wintgen 1972).

4. Vector and Tensor Fields on Manifolds

The fact that manifolds are modelled on the spaces $\mathbf{R}^n$ implies the existence of the tangent space $\mathcal{T}_x(\mathcal{M})$ to a given manifold $\mathcal{M} = (X^n, \{A\})$ at each point $x \in X^n$ that is isomorphic to $\mathbf{R}^n$. (If misunderstanding can be excluded we shall also write $x \in \mathcal{M}$.)

Let us consider two differentiable functions f_1 and f_2 on a manifold $\mathcal{M}$. These functions are said to have the same *germ* at a point $x \in \mathcal{M}$ if there is a neighbourhood of x on which f_1 and f_2 coincide. The *germ of a function* f at x is — from the definition — the equivalence class of functions having the germs at x equal to that of f. Germs of functions at x form a vector space (with respect to addition and multiplication by real numbers). Moreover, since the product of two germs at a point x (defined as the pointwise multiplication of functions) is a germ at x, the germs at x form an algebra.

Finally, let us recall that a linear mapping satisfying the Leibniz rule is called *derivation*. Now, a vector tangent to a manifold $\mathcal{M}$ at a point $x \in \mathcal{M}$ can be defined as a derivation on the algebra of germs of differentiable functions at x. The set of all vectors tangent to $\mathcal{M}$ at x is called the *tangent space* to $\mathcal{M}$ at x and is denoted by $\mathcal{T}_x(\mathcal{M})$. Of course, tangent spaces to a manifold are vector spaces (with respect to addition and multiplication by real numbers).

One could translate the above abstract definitions into a language honoured by the handbook tradition, and say that the tangent vector v_x to a manifold $\mathcal{M}$ at a

point $x \in \mathcal{M}$ is a linear mapping from the space $F(\mathcal{N})$ of differentiable functions, defined on a neighbourhood $\mathcal{N}$ of x, into the space $\mathbf{R}^1$, i.e.,

$$v_x(\alpha f + \beta g) = \alpha v_x(f) + \beta v_x(g),$$

where $\alpha, \beta \in \mathbf{R}^1$ and $f, g \in F(\mathcal{N})$; moreover, the mapping is supposed to satisfy the Leibniz rule

$$v_x(fg) = f(x)v_x(g) + g(x)v_x(f).$$

The set of all such tangent vectors to a manifold $\mathcal{M}$ at $x \in X$ is the tangent space $\mathcal{T}_x(\mathcal{M})$.

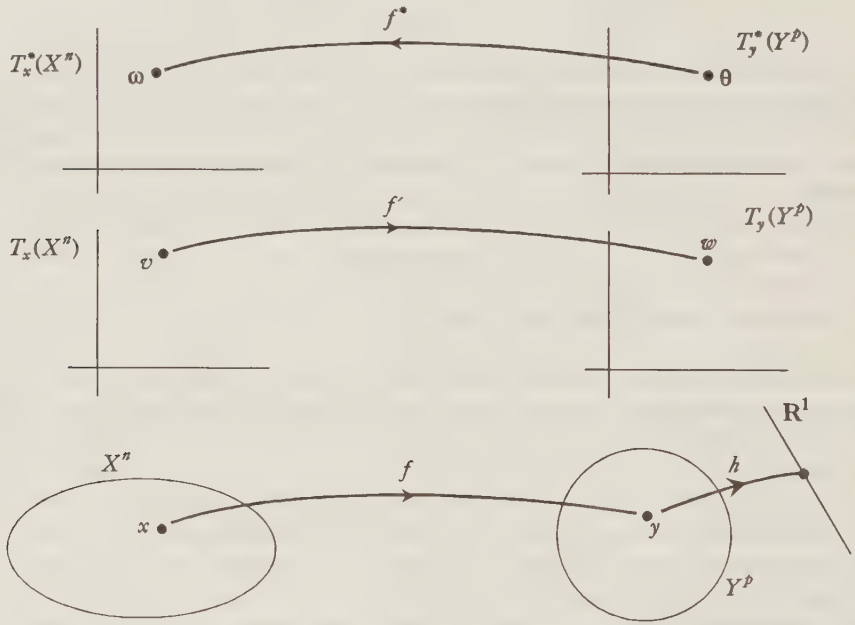


Fig. 1.6. Derivative f' and pull back f^* of a mapping f.

Let us consider the mapping

$$f\colon X^n \longrightarrow Y^p$$

differentiable at $x \in X^n$. The *derivative* of this mapping at x is defined to be a linear mapping

$$f'(x)\colon \mathcal{T}_x(X^n) \to \mathcal{T}_y(Y^p),$$

transforming $v \in \mathcal{T}_x(X^n)$ into $w \in \mathcal{T}_y(Y^p)$ in such a way that, for every function h at the point $y = f(x) \in Y^p$, one has $w(h) = v(h \circ f)$ (Fig. 1.6). It can be easily

seen that under the mapping $f'(x)$ a tangent vector to a curve C in X at a point x is transformed into the tangent vector to the curve $f(C)$ in Y^p at the point y.

Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces. The set $\mathbb{L}(\mathcal{V}, \mathcal{W})$ of all linear mappings from $\mathcal{V}$ into $\mathcal{W}$ is a vector space (with addition and multiplication by scalars, defined in a natural way). The vector space $\mathbb{L}(\mathcal{V}, \mathbf{R}^1)$ is called *dual* to $\mathcal{V}$ and is denoted by $\mathcal{V}^*$. The space $\mathcal{T}_x^*(\mathcal{M})$, dual with respect to tangent space $\mathcal{T}_x(\mathcal{M})$, is called the *cotangent space* to the manifold $\mathcal{M}$ at a point x; its elements are called *covectors*. The following names are also used: *contravariant vectors* — for elements of $\mathcal{T}_x(\mathcal{M})$, and *covariant vectors* — for elements of $\mathcal{T}_x^*(\mathcal{M})$.

Let us consider a differentiable mapping

$$f\colon X^n \longrightarrow Y^p$$

and a covariant vector $\theta \in \mathcal{T}_y^*(Y^p)$. We define the *pull back* of θ (induced by the mapping f) as a linear mapping

$$f^*\colon \mathcal{T}_y^*(Y^p) \longrightarrow \mathcal{T}_x^*(X^n),$$

transforming θ into $\omega \in \mathcal{T}_x^*(\mathcal{M})$ in such a way that, for any $v \in \mathcal{T}_x(X^n)$, the functional equality $(f^*\theta)v = \theta(f'v) \circ f$ is satisfied (or the numerical equality $(f^*\theta)_x v_x = \theta_y(f'v)_y$ (see Fig. 1.6)).

And consequently, the *pull back of a function* g at a point $y = f(x) \in Y^p$ (induced by the mapping f) is defined as $f^*(g) = g \circ f$ (or more precisely: $f^*(g) \mapsto (g \circ f)(x)$).

Let $\mathcal{V}_1, \ldots, \mathcal{V}_n$ and $\mathcal{W}$ be vector spaces. A mapping $f\colon \mathcal{V}_1 \times \cdots \times \mathcal{V}_n \to \mathcal{W}$, linear with respect to each variable, is said to be a *multilinear mapping*. Let $\tau_i \in \mathcal{V}_i$, $i = 1, \ldots, n$. We define the multilinear mapping

$$\tau_1 \otimes \cdots \otimes \tau_n\colon \mathcal{V}_1 \times \cdots \times \mathcal{V}_n \longrightarrow \mathbf{R}^1$$

by

$$\tau_1 \otimes \cdots \otimes \tau_n(v_1, \ldots, v_n) = (\tau_1(v_1)), \ldots, (\tau_n(v_n)),$$

where $v_i \in \mathcal{V}_i$. This function is called the *tensor product*.

4.1. Definition. A real multilinear function

$$f\colon \underbrace{\mathcal{V}^* \times \cdots \times \mathcal{V}^*}_{r \text{ times}} \times \underbrace{\mathcal{V} \times \cdots \times \mathcal{V}}_{s \text{ times}} \longrightarrow \mathbf{R}^1$$

is called *a tensor of the type* (r, s) (over the space $\mathcal{V}$). Tensors of the type (r, s) form the vector space

$$\underbrace{\mathcal{V} \otimes \cdots \otimes \mathcal{V}}_{r \text{ times}} \otimes \underbrace{\mathcal{V}^* \otimes \cdots \otimes \mathcal{V}^*}_{s \text{ times}} = \mathcal{T}^r{}_s(\mathcal{V})$$

(or in the abbreviated form $\otimes^r \mathcal{V} \otimes^s \mathcal{V}^* = \mathcal{T}^r{}_s(\mathcal{V})$). If the dimension of $\mathcal{V}$ is n, $\dim \mathcal{V} = n$, then $\dim \mathcal{T}^r{}_s(\mathcal{V}) = n^{r+s}$. Tensors of the type $(0, 0)$ are scalars, tensors of

the type $(1,0)$ are contravariant vectors, and tensors of the type $(0,1)$ are covariant vectors.

A tensor of the type (r,s) *on a manifold* $\mathcal{M}$ is defined to be a real multilinear function on

$$\otimes^r \mathcal{T}_x(\mathcal{M}) \otimes^s \mathcal{T}_x^*(\mathcal{M}) = \mathcal{T}^r{}_s(\mathcal{T}_x(\mathcal{M})).$$

A tensor field $\mathcal{T}$ *of the type* (r,s) *on a manifold* $\mathcal{M}$ is defined to be a mapping $\mathcal{M} \to \mathcal{T}^r{}_s(\mathcal{M})$ such that, for every $x \in \mathcal{M}$, one has $\mathcal{T}(x) \in \mathcal{T}^r{}_s(\mathcal{T}_x(\mathcal{M}))$. A tensor field of the type $(1,0)$ on $\mathcal{M}$ is a field of contravariant vectors on $\mathcal{M}$; a tensor field of the type $(0,1)$ (of the class $\mathcal{C}^k$) is also called a *one-form*. A tensor field is called *symmetric* (*antisymmetric* or *skew-symmetric*) if, at each point $x \in \mathcal{M}$, $\mathcal{T}(x)$ is a symmetric (antisymmetric) tensor. Totally antisymmetric tensor field of the type $(0,p)$ (of the class $\mathcal{C}^k$) is called a *p-form*.

The pull back mapping can be generalized to any p-form. Let f be a differentiable mapping

$$f\colon\ X^n \to Y^q,$$

and ω any p-form on Y^q. The *pull back* $f^*\omega$ of the p-form ω (induced by the mapping f) is defined in the following way,

$$f^*(\omega)(v_1,\ldots,v_p) = \omega(f'v_1,\ldots,f'v_p),$$

where $v_i \in \mathcal{T}_x(X^n)$, and $f'v_i \in \mathcal{T}_{f(x)}(Y^q)$.

5. Submanifolds

Let $f\colon \mathcal{N}^q \to \mathcal{M}^p$ be a differentiable mapping from a differential manifold $\mathcal{N}^q$ to a differential manifold $\mathcal{M}^p$. This mapping is said to be of *rank* r at a point $y \in \mathcal{N}^q$ if the dimension of $f'(\mathcal{T}_y(\mathcal{N}^q))$ is r. The following, almost self-evident, relationships hold:

if the rank of f at y is q $(= \dim \mathcal{N}^q)$, the induced mapping f' at y is injective, and $q \leq p$;

if the rank of f at y is p $(= \dim \mathcal{M}^p)$, the mapping f' at y is surjective, and $q \geq p$.

If the rank of the mapping f (of the class $\mathcal{C}^k$) at each point of the manifold $\mathcal{N}^q$ is q, f is called the *immersion* (of the class $\mathcal{C}^k$), and the manifold $\mathcal{N}^q$ is said to be *immersed* in the manifold $\mathcal{M}^p$. Let us note that the set $f(\mathcal{N}^q)$ need not carry the manifold structure since f need not be injective (however, if f is an immersion, f' is always an injection; this immediately follows from the above statements and definitions). Immersion that is also an injection is called *embedding*, and one correspondingly speaks of an *embedded* manifold into one another. If f is an embedding, the set $f(\mathcal{N}^q) \subset \mathcal{M}^p$ with the differential structure induced by f is a manifold, although this differential structure need not be consistent with the original differential structure on

$\mathcal{M}^p$; if it is consistent, f is said to be a *regular embedding*. If f is a regular embedding, $f(\mathcal{N}^p)$ is called[3] a *submanifold* of $\mathcal{M}^p$.

An open subset of a manifold $\mathcal{M}^p$ considered, in a natural way, as a submanifold of $\mathcal{M}^p$, is called an *open submanifold* of $\mathcal{M}^p$.

6. Manifold as a Model of Space–Time

All macroscopic theories of contemporary physics — and theories that have played important roles in the history of physics (provided they have been reconstructed in the language of modern mathematics, see Appendix) — presuppose that the arena in which physical processes occur has the structure of a four–dimensional differential manifold. Points of this manifold are called *events*. Events are determined by their four coordinates in local charts. Local charts are treated as local reference frames (see below, sec. 7, A). One of the coordinates in each of such charts is considered to be the time coordinate, the remaining three — space coordinates. This distinction of time and space coordinates has a purely local meaning: it is defined only in a local chart. One should expect that the possibility to cover the entire manifold with a single chart (i.e. the possibility of an atlas consisting of one chart, as it is the case in $\mathbf{R}^n$) is a rare exception. Since from the mathematical point of view time and space coordinates are on equal footing, one often prefers speaking of four coordinates of events or — to be more precise — saying that a four–dimensional differential manifold serves as a *mathematical model of the physical space–time*. With less care, but more easy, one says simply that the space–time is a four–dimensional manifold. Therefore, it can be asserted that space–time is a set of all possible events, where by an event one should mean an "idealized occurrence in the physical world having extension neither in space nor in time" (Geroch 1978, p. 3), for instance, an idealization of an elastic collision of two particles.

The four–dimensionality of space–time is an empirical fact. A physical counterpart of the dimension is a number of the degrees of freedom of a physical system. For instance, three degrees of freedom in a translational motion correspond to the three dimensionality of space. The possibility to move along only one temporal direction is reflected in the one dimensionality of time. Let us notice, however, that the "arrow" along this direction (the so–called *arrow of time*) has not been determined by the model. The fact (known from the very personal experience) that physical processes can "move" only from the past to the future has no explanation within the present model. At most, we can, in a purely conventional manner, call one end of the temporal direction the *future*, and the other and — the *past*. In the following, we shall often make use of this convention.

[3]One should note the difference in terminology used by different authors (see, for example, Choquet–Bruhat et al. 1982, pp. 239–242; Bishop and Goldberg 1980, pp. 40–43; Kobayashi and Nomizu 1963, chapt. 1, sec. 1; Brickell and Clark 1970, pp. 68–83; O'Neill 1983, pp. 15–21; Hawking and Ellis 1973, pp. 23–24).

Usually, it is assumed that a manifold modelling the physical space–time is orientable, and, additionally, that it is separately time and space orientable (this will be defined in chap. 2, sec. 1; see also below sec. 7, B, and sec. 1. 3 in Sachs and Wu 1977).

The manifold model ascribes to space–time not only the continuity but also (in general) a "sufficiently smooth" continuity. Usually, it is assumed that all mappings appearing in the definition are of class $\mathcal{C}^k$, where k is as large a number as it is required by a given physical situation. Most often, one simply assumes[4] that a modelling manifold is of class $\mathcal{C}^\infty$.

The postulate of continuity leads to the question of the status of this concept in microphysics. According to quite a common opinion, it is a smooth space–time manifold that continues to play the role of an arena for physical processes in quantum mechanics. However, one could quote convincing reasons supporting the view that a smooth space–time is a macroscopic concept: we coin this concept within the world of macrophysics, in which all measurements are performed providing us with information about the micro–world, and then "projecting it down" into the domain of quantum phenomena. Such a view is corroborated by an analysis of the mathematical structure of quantum mechanics. Unfortunately, this problem goes beyond the limits of the present book: in the following — except for sporadic remarks — we will deal only with space–times of macroscopic physical theories which are always modelled by smooth manifolds.

Finally, one should note an interesting methodological problem: can the manifold model of space–time be defined operationally (or at least quasi–operationally)? And, if the answer is positive, how can it be defined? According to Trautman's (1964) recipe, a space–time manifold can be constructed with the help of "negligibly small" clocks, the relative positions of which are parametrised by triples of real numbers. In the construction proposed by Synge (1965), a fundamental tool is again a clock, and a fundamental operation — a time measurement. However, this author tacitly assumed the existence of signals that would inform the observer on distant events. A construction, free from such gaps, was proposed by Ehlers, Pirani and Schild (1972). In their approach, free fall and light propagation serve as "elementary operations" (see also Ehlers 1973). These operations allow one to construct not only the manifold, but also the metric structure of space–time (see, next chapter and Heller 1977).

7. Comments and Remarks

A. Local charts on a manifold correspond to local frames of reference in space–time (although not all coordinate transformations a physicist would consider as changing to new frames of reference (see, Dixon 1978, pp. 2–7). Full discusion of this problem

[4]Some authors, when speaking of a $\mathcal{C}^k$ space–time, have in mind the fact that the metric defined on this space–time is of class $\mathcal{C}^4$ (see, for instance, Tipler et al. 1980, pp. 155–157).

is possible only after a metric has been introduced on a manifold (see, chapter 2, sec. 5, D). The fact that a manifold admits — on the strength of its definition — all atlases equivalent to a given atlas (see above, sec. 2) shows a certain "relativity" of the space–time model discussed here: the very concept of space–time includes all possible classes of reference frames consistent with it, and none of these classes is *a priori* distinguished in any way. Some classes of reference frames may become privileged only afterwards, when the space–time considered is enriched by additional postulates which should always be physically motivated.

B. The problem of space–time orientability, space orientability and time orientability belongs to the most fundamental problems of physics and cosmology. Having in mind that at the present stage of our analysis we have no "generally valid" method to split space–time into space and time separately, let us refer to interesting remarks of a mathematician, Laurent Schwartz (1967, vol. 2, chapt. 6, sec. 5), who after asking the question

"Is our universe an orientable manifold?" continues: Let us put aside the relativistic point of view that teaches us to look at the universe as a four–dimensional manifold, and let us assume that our world is modelled by a three–dimensional $\mathcal{C}^\infty$ manifold. Is it orientable? Do the known laws of physics allow us to equip it with a canonical orientation? To begin, let us assume that our world, considered as a three–dimensional manifold, is not orientable. This means that in the Universe there exist some paths, similar to those in the Möbius strip, such that starting a journey along one of them with a certain orientation, and pursuing the orientation continously during the travel, one comes back to the staring point with the opposite orientation. A human being travelling throughout the Universe along such a path, after coming back to the Earth, would find that its heart is at the right side of its body, and its liver at the left side. Letters in a book, read during the travel, would now run from right to left. All this without any sudden change experienced by the traveller during his trip. Anyway, after his coming back he would judge himself absolutely normal and "invariant". On the other hand, he would think that all phenomena on the Earth had been "reversed" during his absence. Schwartz concludes that all arguments for the world's orientability, quoted in elementary textbooks of physics and mathematics, rest on a very weak basis.

C. From the times of Einstein, there exists in physics a set of ideas called the *doctrine* or the *principle of Mach*. These ideas go back to Leibniz and Berkeley; however, they were read out by Einstein from the works of Ernst Mach in which this author criticised the Newtonian mechanics. Roughly speaking, Mach's principle postulates a causal relationship between the local physics and the global structure of space–time (for an account of Mach's original doctrine, see Heller 1988). In its more maximalistic formulations Mach's principle requires that the global structure of the world should *uniquely* determine the local physics[5]. (For more details see Goenner 1970, Reinhardt 1973, Heller 1975a, Raine and Heller 1981, Raine 1981a, Barbour 1983). As the number of quoted works testifies (and this is only an "unfair sample" of the continuously increasing literature), discussions on the subject are still going on. It seems that the status of Mach's principle in the theory of general relativity

[5]In the original Einstein's formulation Mach's principle postulates the "relativity of mass": the mass of a test body should be uniquely determined by the distribution and motions of all masses in the Universe.

has been clarified in the works of Raine (1975, 1981a), but Mach's ideology manifests some tendencies to go beyond this theory and to provide new inspirations in creating new physical theories (see, for example, Barbour 1974, 1983; Barbour and Bertotti 1977; Raine and Heller, chapter 13). I do not want now to go deeper into this set of problems; I would like only to note that the manifold model of space–time imposes substantial restrictions upon the possibilities to implement Mach's doctrine into a physical theory. Any manifold has — from its very definition — local topological and differential properties identical with those of the modelling space $\mathbf{R}^n$. Although this circumstance leaves a certain freedom to manipulate the local structure of space–time with the help of field quations (see, Raine 1975, 1981a), this does not change the fact that space–time must locally be $\mathbf{R}^n$ irrespectively of its global structure. This limitation can be removed only by postponing modeling space–time by a differental manifold (see, below, comment E).

D. The manifold concept has been defined in terms of the modelling space $\mathbf{R}^n$. A manifold can be also constructed by suitably gluing copies of $\mathbf{R}^n$ (see, above, sec. 2). In this sense, the arithmetical space $\mathbf{R}^n$ is *a priori* with respect to the manifold concept. However, it is not a logical necessity, but first the result of our decision on what a manifold is, and second the question of agreement of this decision with empirical data. Under the present precision of empirical data the model can, in principle, be discarded and replaced by a new one. Kant was wrong when asserting that the Euclidean character of space is an *a priori* condition of any experiment.

E. A host of empirical data testifies that, within the domain controlled by macrophysics, there is no need to give up the manifold model of space–time, although it is not impossible — as we have seen — that, in microphysical theories, space–time appears from outside, as it were, namely as a consequence of the fact that an observer performs all his measurements being immersed in the macroscopic space–time. The necessity to go beyond the manifold model of space–time could appear within a domain of quantum gravity. There are strong reasons compelling one to believe that a quantisation of gravity — first of all, at very early stages of the world's evolution, in the vicinity of the initial singularity — could hardly be avoided. Although some other proposals are known, it seems highly probable that the quantisation procedure would require giving up (or rather to generalize) the manifold model of space–time. As an example of this view, let us quote from Trautman's work:

> "The topological and differential structures of space–time do not seem to possess a well–defined operational meaning. Therefore, it is likely that they will have to be abandoned, or rather replaced by another structure which will be more closely related to, and influenced by, physical phenomena than the absolute, locally Euclidean manifold structure of space–time assumed in all current theories. In my opinion, a satisfactory quantum theory of space, time and gravitation will have to do away with the notion of a differentiable manifold as a model of space–time" (Trautman 1973b, p. 183).

Of course, one should expect that the future, long–expected quantum theory of gravity would give, as its limiting case, the present manifold model of space–time. For this reason this model ought to be regarded as a permanent achievement of

contemporary physics.

Bibliographical Note. Fundamental information on Banach spaces could be found in the two–volume book of Maurin (1976, 1980), a more extensive study of functional analysis in the classical monograph by Riesz and Nagy (1972), and a very transparent introduction in the book by Roman (1975, vol. 2). Chapter 7A of the book by Choquet–Bruhat et al. (1982) gives a short account of differential manifolds modelled on infinitely–dimensional topological spaces.

The concept of (finitely–dimensional) differential manifold is presented in all handbooks of modern differential geometry. From among many excellent books in this field, the two volumes by Kobayashi and Nomizu (1963) merit our special attention. This book is often quoted by physicists, and many works of differential geometry written after 1963 are based on it and often contain generalizations or applications of its results. More concise, but also very good, accounts of differential geometry can be found in Bishop and Crittenden (1964) or Hicks (1965). The books by Auslander and MacKenzie (1963), and by Brickell and Clark (1970) are devoted entirely to manifolds; the last one contains many useful examples. The reader interested in an introductory exposition should consult Wallace (1968) or O'Neill (1966). The first of these books deals with a topological aspect of manifolds, the second is an elementary, but modern, exposition of differential geometry with a number of computational exercises; the manifold concept is introduced gradually as a generalization of the concept of plane in the Euclidean space. The book by Schutz (1984) is especially recommended because of its didactic values. The reader is introduced to both modern geometric structures and computational techniques. The selection of the material has been made in view of future applications in physics; such applications are also discussed. Conceptual foundations of the manifold theory are very clearly discussed in a little book by Isham (1989).

The textbook of differential geometry written by Sikorski (1972) is exceptional in a certain sense. The exposition of differential geometry is here based on a concept of differential space, introduced by the author. A *differential space* is defined to be the pair $(\mathcal{M}, \mathcal{C})$ where $\mathcal{M}$ is any set, and $\mathcal{C}$ a family of real functions on $\mathcal{M}$, satisfying two conditions that axiomatically define the "smoothness" of functions belonging to $\mathcal{C}$, namely: (1) $\mathcal{C}$ is closed with respect to localization, (2) $\mathcal{C}$ is closed with respect to superposition with Euclidean functions. To be more precise, let $(\mathcal{M}, \tau)$ be any topological space, and $\mathcal{C}$ any non–empty family of functions on $\mathcal{M}$. A function f, defined on $\mathcal{N} \subset \mathcal{M}$, is said to be a *local $\mathcal{C}$–function* if, for every $p \in \mathcal{N}$, there is a neighbourhood $\mathcal{U}$ in the topological space $(\mathcal{N}, \tau_N)$, where τ_N is a topology induced in $\mathcal{N}$ by τ, and a function $g \in \mathcal{C}$ such that $f|_U = g|_U$. The set of local $\mathcal{C}$–functions is denoted by $\mathcal{C}_A$; one obviously has $\mathcal{C} \subset \mathcal{C}_M$. The family of functions $\mathcal{C}$ is said to be *closed with respect to localization* if $\mathcal{C} = \mathcal{C}_M$. Let, further, $\mathcal{E}$ be the set of all $\mathcal{C}^\infty$ functions defined on $\mathbf{R}^n$ called *smooth Euclidean functions*. $\mathcal{C}$ is said to be *closed with respect to superposition with smooth Euclidean functions* if, for any function $\omega \in \mathcal{E}$ and any $n \in \mathbf{N}$, the fact that $f_1, \ldots, f_n \in \mathcal{C}$ implies that $\omega \circ (f_1, \ldots, f_n) \in C$. The differential space concept is a vast generalization of the manifold concept. *Every* subset of $\mathbf{R}^n$ is a differential space, but there exist many differential spaces which can be embedded in no $\mathbf{R}^n$ (even if $n = \infty$). In particular, each differential manifold is a differential space. A diffeomorphism between two differential spaces can be defined in a natural way. By postulating that a differential space is locally diffeomorphic to a Euclidean space, one recovers the concept of a differential manifold. It turns out that, in spite of its generality, differential space is a workable concept. In particular, one can meaningfully define a tangent space to a differential space at any of its points. Large parts of differential geometry can be developed on differential spaces proper (i.e. on such differential spaces that are not manifolds). It is worthwhile to notice that Walczak and Waliszewski (1981) have published a collection of exercises (in general, difficult), especially adapted to Sikorski's book. Penrose and Rindler (1984) define space–time manifold as a pair $(\mathcal{M}, \mathcal{C})$, without however, generalizing it to space–time differential spaces.

In Gruszczak et al. (1988, 1989) a proposal has been worked out to model space–time by a differential space rather than by a differential manifold. Such a generalized space–time model can find its applications in a quantum gravity regime (when space time is still a differential space

but already not a differential manifold), and seems to be very promising in describing space–time singularities. Unfortunately, Sikorski (1972), and Walczak and Waliszewski (1981) exist only in Polish. For a foreign reader an extensive review by Heller et al. (1989) could be useful of differential space methods and its possible applications to space–time geometry.

There exist many expositions of differential geometry, written especially for physicists. Those by Choquet–Bruhat et al. (1982), Bishop and Goldberg (1980), and O'Neill (1983) are the best for our purposes. Useful information concerning the differential structure can be found in the second volume of the book by Richtmyer (1981, chapters 23–28). The reader interested in a short review of modern methods in differential geometry should consult a paper by Schmidt (1972) or introductory chapters of such relativity theory textbooks as: Misner et al. (1973), Schutz (1985), Ryan and Shepley (1975); the last one is on a more advanced level than the first two. Thirring (1978) contains a transparent presentation of elementary concepts of the manifold theory (with applications to classical physics) and a number of instructive exercises.

The manifold model of space–time has been submitted to a detailed analysis by Hawking and Ellis (1973), and Trautman (1964, 1968, 1970a, 1973b, 1976, 1978). See also the book by Raine and Heller (1981), and the works of the present author (Heller 1981a, 1986).

Chapter 2

Relativistic Model of Space–Time

0. Introduction

The space-time model introduced in the previous chapter is very general. To this property it owes its applications in all major theories of macroscopic physics. The manifold structure ensures, among others, the existence of local coordinate systems on a manifold, and the possibility to do calculus on it. Because of that, concepts such as: position, velocity and acceleration are well determined (at least locally) in all macroscopic space–times. However, the manifold model is too poor to allow the full physics to "develop" on it. In particular, such important concepts as space and time measurements cannot be defined within the conceptual framework presupposed by this model. Space and time measurements become meaningful only after introducing a *metric* on a manifold (or on certain of its submanifolds, as it takes place, for instance, in the case of space–time of classical mechanics; see Appendix). The remark in parentheses suggests that various physical theories will differ from each other in the way they introduce a metric structure on a space–time manifold. In the present chapter, we will analyse the metric structure demanded by the relativistic theories of space–time, i.e, by special and general relativity theories, that is to say the *Lorentz metric structure*.

The question "what are the (necessary and sufficient) conditions of the existence of the Lorentz metric on a differential manifold?" is of uppermost importance for theoretical physics. If we agree with Maxwell's saying that in physics one does not know of which one speaks unless one has a recipe *how to measure*, we must admit that there is no meaning in speaking of *space-time physics* as long as there is no metric defined on a corresponding space–time manifold. Therefore, the question of the Lorentz metric existence is, in a sense, the question of the possibility of physics. The metric structure is very rich. It contains several other structures (conformal, projective, affine, ...). These structures can be introduced axiomatically starting from the bare topological and differential structures. In this way, by gradually enriching the construction with new properties, ensured by subsequent axioms, one finally obtains the full Lorentz manifold (see Ehlers et al. 1972). However, in this chapter we will adopt a different strategy: we introduce a "ready–made" Lorentz metric on a manifold which implicitly contains all other afore–mentioned structures; they will be used and, if necessary, further analysed in subsequent parts of the present book.

Mathematical richness of the Lorentz structure is paralleled by its paramount importance in physics. As it is expressed by Sachs and Wu (1977b, p. 12):

"[The Lorentz metric] g somehow remembers the right things and forgets the wrong ones. Concepts of causality, distance, time, velocity, speed, acceleration, rotation, rigidity, simultaneity,

orthogonality, gravity, and so on, are derived from g to the extent they are retained at all. g therefore must play many roles; its unifying power is remarkable."

In the present chapter, the organization of the material is the following. Section 1 introduces a metric on a manifold, changing it into a metric manifold, and sec. 2 gives a physical interpretation to this procedure. Section 3 presents, in a more detailed manner, a special case of a metric space–time model, namely the Minkowski space–time which constitutes a geometric arena for the special theory of relativity. Finally, in sec. 4, necessary and sufficient conditions are given for the existence of a Lorentz metric on a differential manifold. Besides the standard treatment of this problem, its formulation in the language of cohomology groups is briefly presented. In *Comments and Remarks* (sec. 5) we elucidate a relationship between the existence of the Lorentz structure on a manifold and the concept of time. Results obtained in this chapter seem to suggest that a certain form of temporality is a necessary condition for physics to happen on a manifold.

1. Metric Manifolds

1.1. Definition. Let X be a differential manifold. A $\mathcal{C}^k$ *metric* (or $\mathcal{C}^k$ *metric field*) on X is a $\mathcal{C}^k$ tensor field of the type $(0,2)$ on X, satisfying the following conditions:

(1) g is symmetric, i.e., for every $x \in X$, the tensor g_x (g at a point x) is a symmetric tensor;

(2) for every $x \in X$, the bilinear form $g_x(.\ ,\ .)$ is non–degenerate, i.e., $g_x(v,w) = 0$, $v, w \in \mathcal{T}_x(X)$, for all v, if and only if $w = 0$.

One speaks also of the *metric tensor* g. A differential manifold with a metric, defined in this way, is said to be a *Riemann manifold* (or a *Riemannian manifold*). Sometimes one says that a manifold X is equipped with a *Riemann structure* (or a *Riemannian structure*). A Riemann manifold is said to be a *proper Riemannian manifold* if, for every $0 \neq v \in \mathcal{T}_x(X)$, $x \in X$, one has $g_x(v,v) > 0$. If this is not the case, one speaks of a *pseudoriemannian manifold.*

A metric tensor g equips the tangent vector spaces $\mathcal{T}_x(X)$, $x \in X$, with scalar products $g_x(v,w) = g_{ij}v^iw^j$, $v, w \in \mathcal{T}_x(X)$. As it is well–known, in the Euclidean space, with the help of the scalar product, one can measure lengths of vectors and angles between vectors. These operations are naturally transferred to Riemannian manifolds. In particular, the norm ("length") $\|v\|$ of a vector $v \in \mathcal{T}_x(X)$ is defined as $\|v\|^2 = g_x(v,v) = g_{ij}v^iv^j$. In pseudoriemannian manifolds, it can be $\|v\| = 0$, for $v \neq 0$; such a vector v is called *null* (or *lightlike*).

With a suitable choice of a basis (e_i) for $\mathcal{T}_x(X)$ the quadratic form $q = g_x(v,v)$ can be written as

$$q = g_{ij}v^iv^j = -\sum_{i=1}^{k}(v^i)^2 + \sum_{i=k+1}^{n}(v^i)^2.$$

The quadratic form q is often denoted by ds^2 and is also called a *line element* of the manifold $\mathcal{M}$. The basis, in which the quadratic form $g_x(v,v)$ assumes the above

form, will be called *pseudoorthogonal*. The number k, often denoted by I also, is called the *index* of q; and the number $(n - I) - I$ its *signature*.[1] Because of the metric continuity, the index and signature are the same for all $x \in X$, provided the manifold X is connected. If $I = 0$ (or $I = n$), the manifold is of course a proper Riemannian manifold.

1.2. Definition. A manifold for which $I = 1$ (or $I = n - 1$) is said to be a *Lorentz manifold* (*Lorentzian* or sometimes *hyperbolic*). Vector spaces $\mathcal{T}_x(X)$ tangent to Lorentz manifolds are said to be *Minkowski manifolds*. Correspondingly, one speaks of *Lorentz* or *Minkowski metrics*.

All non–zero vectors $v \in \mathcal{T}(X)$, where X is a Lorentz manifold, are classified (independently of the choice of a basis) into *timelike*, *null*, and *spacelike* according to whether $g_x(v, v)$ is less than, equal to, or larger than zero. The set of all null vectors $L_x \in \mathcal{T}_x(X)$ is called a *light cone* at x. The light cone L_x separates all timelike vectors in $\mathcal{T}_x(X)$ from all spacelike vectors in $\mathcal{T}_x(X)$. The set T_x of all timelike vectors in $\mathcal{T}_x(X)$ is an open set in $\mathcal{T}_x(X)$ and has two connected components, T_x^+ and T_x^-, which are **conventionally** called the *future* and the *past* of x. Analogously, the light cone L_x has two connected components, L_x^+ and L_x^-, which are called the *future light cone* and the *past light cone* of x, respectively.

The set $T \equiv \bigcup_x T_x$, $x \in X$, of all timelike vectors in a Lorentz manifold X can have one or two connected components (see Sachs and Wu 1977b, 1.2). Let X be a connected Lorentz manifold. If the set T has two components, X is said to be *time orientable*; if T has one component, X is said to be *time nonorientable*. If one of the two components of T is defined to be *future oriented*, and the remaining one as *past oriented*, the manifold X is called *time oriented*. One should notice that the geometry itself does not distinguish any of the two possible time orientations. At this stage of our analysis the choice of one of them is a matter of convention.

Let X^n and Y^n be differential manifolds with metrics g_1 and g_2, correspondingly. A mapping $f: X^n \to Y^n$, which is a diffeomorphism and preserves the metric, i.e., $g_1 = f^* g_2$, is said to be an *isometric mapping* or shortly an *isometry*; in such a case, the manifolds X^n and Y^n are said to be *isometric*. A Riemann manifold, isometric with the manifold $(\mathbf{R}^n, \mathrm{id})$ with the metric $g = \sum_{i=1}^n \varepsilon_i dx^i dx^i$, where $\varepsilon_i = \pm 1$, is called a *flat space*. If $\varepsilon = +1$, for every i, the manifold is called a *Euclidean space* and is usually denoted by $\mathcal{E}^n$. Minkowski spaces have the metric

$$\eta = -dx^0 dx^0 + \sum_{i=1}^{i-1} dx^i dx^i,$$

therefore, they are *flat spaces*. The bilinear form η is called a *Minkowski metric*.

[1] One should notice different definitions of index and signature by various authors, and also different conventions concerning signs of some geometric formulae. A list of these conventions can be found on the inner side of the cover of the textbook by Misner et al. (1973).

2. Metric Model of Space–Time

Differential manifolds equipped with a metric structure model spaces or space–times of many macroscopic physical theories. For example, instantaneous spaces of classical mechanics are proper Riemann manifolds, more precisely, three–dimensional Euclidean spaces. One should note, however, that space–time of classical mechanics is not a metric manifold, since there is no meaningful operation in it allowing one to measure space–time intervals between pairs of non–simultaneous events (see Raine and Heller 1981, chapt. 7, and Appendix at the end of the present book). Physical spaces modelled by metric manifolds need not be Euclidean spaces. For instance, instantaneous spaces of the Friedman–Lemaître cosmological models are three–dimensional, not necessarily flat, Riemann manifolds (see, e.g., Bondi 1965, Weinberg 1972, Sciama 1975, Raine 1981b).

Successes of special and general relativity have convincingly demonstrated the efficiency of Lorentz manifolds in modelling the physical space–time. The Lorentz structure is much more subtle than the proper Riemann structure, and it turns out that many important physical results are but almost trivial conclusions following directly from the Lorentz structure.[2] With no exaggeration one can say that four–dimensional manifolds with a Lorentz structure play the role of a standard model of the macroscopic space–time in contemporary physics. In the rest of this book we shall exploit this model. Now, we make a convention that — unless something to the contrary is directly expressed — by a physical model of space–time we shall understand the pair $(\mathcal{M}, g)$, where $\mathcal{M}$ is a four–dimensional $\mathcal{C}^k$ manifold with respect to which all remarks of chapter 1, sec. 6 are valid, and g is a $\mathcal{C}^r$ Lorentz metric, with the $+2$ signature, defined on $\mathcal{M}$.[3] We shall assume — unless directly expressed otherwise — that both the manifold $\mathcal{M}$ itself and the Lorentz metric g, on it are continuously differentiable as many times as it is necessary; in practice, it means that $\mathcal{M}$ and g are of class $\mathcal{C}^\infty$. The pair $(\mathcal{M}, g)$ will be called a *relativistic model of space–time*.

Now, some specifics must be introduced. As it can be easily checked, any isometry is an equivalence relation and, strictly speaking, it is not a single pair $(\mathcal{M}, g)$ but an equivalence class of such pairs that is a model of space–time; two pairs $(\mathcal{M}_1, g_1)$ and $(\mathcal{M}_2, g_2)$ are equivalent, if they are isometric, i.e. if there is a mapping $f: \mathcal{M}_1 \to \mathcal{M}_2$ such that $f^* g_2 = g_1$. A metric model of space–time does not distinguish between two manifolds if they are isometric. We should keep this in mind, although one usually works only with one pair $(\mathcal{M}, g)$ treating it, in a sense, as a representative of the entire equivalence class (see Hawking and Ellis, p. 56). Traditional physics is interested in local properties of space–time; cosmology, from its very nature, contemplates global properties of space–time. And here we are met with a subtle, but rich in consequences,

[2] One can see it clearly in the book by Sachs and Wu (1977b) aimed at presenting general relativity theory in the form of a mathematical text, as far as it is possible. In consequence of this endeavour significant parts of the theory changed into concise corollaries of mathematical theorems.

[3] From the manifold definition it follows that $r \leq k$.

difficulty: how to guarantee that we are dealing with the "whole of space–time", that no part of if has been "artificially" cut off? First of all, one must make more precise intuitions contained in this question. The following definitions serve this end. A space–time $(\mathcal{M}', g')$ is said to be a $\mathcal{C}^r$ *extension* of a space–time $(\mathcal{M}, g)$ if there is an isometric $\mathcal{C}^r$ embedding of the manifold $\mathcal{M}$ into the manifold $\mathcal{M}'$, $f: \mathcal{M} \to \mathcal{M}'$. A space–time is said to be $\mathcal{C}^r$ *inextensible* if, for every its isometric $\mathcal{C}^r$ embedding, $M \to M'$, one has $f(M) = M'$, i.e. space–time is inextensible if it is identical with every its extension (see Hawking and Ellis 1973, pp. 58–59). Geroch (1970) and Clarke (1976) have shown that every Lorentz manifold can be extended as long as it becomes inextensible.

Inextensibility is a mathematical counterpart of the intuitive concept of "totality" of space–time. In the following, we shall assume — unless directly stated otherwise — that all space–times we shall be dealing with are inextensible in a suitable class of smoothness. It is evident that the specification of this class of smoothness is crucial. If we diminish smoothness requirements, an inextensible space–time may become extensible. The question how to suitably choose the smoothness condition has far–reaching consequences in cosmology (see Tipler et al. 1980, pp. 155–158).

Usually, one ascribes some other properties to a metric manifold as a space–time model (besides its inextensibility), for instance: orientability, orientation (see chapt. 1, secs. 6 and 7B), time orientability, etc. (see Sachs and Wu 1977b, sec. 1.3). At the present state of our analysis we shall not assume such additional postulates; they will be discussed and, if necessary assumed, later on along with development of our presentation.

Only after a metric model of space–time has been established, the question becomes meaningful: which are the *largest* and the *smallest* scales for which this models preserves its validity? The question about the upper limit of space–time scales is clearly connected with the cosmological extrapolation (in space and time); this problem will inevitably come back in the following pages of the present book. The question about the lower limit of space–time scales is one of the key problems of the present subquantum physics (see remarks in chapt. 1, sec. 7, E). In the view of Hawking and Ellis (1973, p. 57), there is empirical evidence that space–time has a structure of a smooth manifold down to the scales of 10^{-15} cm. The standard cosmological model suggests that this limit can be further lowered down to the Planck length, i.e., to scales 10^{-33} cm (see, for instance, Hawking and Ellis 1973, pp. 362–364; Misner et al. 1973, pp. 1196–1203; Demiański 1985, pp. 281–283). Doubtlessly, these problems belong to the most fascinating questions of contemporary physics.

3. Space–Time of Special Relativity

The pair $(\mathcal{M}, \eta)$, where $\mathcal{M}$ is a four dimensional manifold and η the Minkowski metric on $\mathcal{M}$ (see, sec. 1), is a space–time model in special relativity. This theory was created by Einstein in 1905, and it was Herman Minkowski to whom the theory owes its transparent geometric form (Minkowski 1907, 1908a, b; the last work is often quoted).

There is a basis $(e_\alpha) = (e_0, e_k)$ in which the components of the Minkowski metric assume the form $\eta_{\alpha\beta} = \mathrm{diag}(-1, 1, 1, 1)$. This basis will be called *pseudoorthogonal*. The coordinate with the index zero is interpreted as time coordinate, the remaining ones as space coordinates?[4]

The group of linear transformations

$$\mathcal{P}\colon\ (e'_\beta) \longmapsto a^\alpha{}_\beta e_\alpha + b^\beta,$$

where $a^\alpha{}_\beta$ and b^β are constants, such that

$$a^\alpha{}_\mu \eta_{\alpha\beta} a^\beta{}_\nu = \eta_{\mu\nu},$$

that is to say, preserving the bilinear form $\eta_{\alpha\beta} v^\alpha v^\beta$, is called a *Poincaré group* or *non–homogeneous Lorentz group* and is usually denoted by $\mathcal{P}(4)$. If $b^\beta = 0$, the name a *Lorentz* or *homogeneous Lorentz group* is used; it is denoted by $\mathcal{L}(4)$.

It can be shown that $\det[a^\alpha{}_\beta] = \pm 1$. If $a^0{}_0 > 0$, the transformations, being elements of the Lorentz (homogeneous) group, form the so–called *ortochronous Lorentz subgroup*; they preserve the time orientation (i.e., the direction of timelike and null vectors).

If $a^0{}_0 > 0$ and $\det[a^\alpha{}_\beta] = +1$, one obtains α subgroup, called a *proper Lorentz group* which is denoted by $\mathcal{L}_0$.

If $a^0{}_0 > 0$ and $\det[a^\alpha{}_\beta] = -1$, the transformations contain time reflections

$$t\colon\ (e_0, e_k) \longmapsto L(-e_0, e_k)$$

and space reflections

$$s\colon\ (e_0, e_k) \longmapsto L(e_0, -e_k);$$

they do not form a subgroup (however, they do form a subgroup together with transformations for which $\det[a^\alpha{}_\beta] = +1$).

The full Lorentz group $\mathcal{L}(4)$ splits into four components: $\mathcal{L}_0$, $t\mathcal{L}_0$, $s\mathcal{L}_0$, and $st\mathcal{L}_0$ (see Choquet–Bruhat et al. 1982, p. 290; Carmeli 1977, chapt. 2).

Special relativity is in fact a theory of invariants of the Lorentz group (or of the Poincaré group) together with a physical interpretation adjoined to it. The classification of vectors into timelike, null and spacelike, and consequently null cones at each point of a space–time manifold (the so–called *cone structure of space–time*) is invariant with respect to transformations forming the Lorentz group. Curves to which the tangent vectors at all their points are timelike, null, or spacelike are called *timelike, null (lightlike)*, or *spacelike*, respectively. In the physical interpretation, timelike curves are treated as representing histories of particles (with a non–zero rest–mass) or of observers, and null curves as representing histories of photons (or more generally of zero rest–mass particles). Spacelike curves, in principle, could represent histories

[4]From now on we will stick to the following convention: Greek indices assume 0, 1, 2, 3 values, whereas Latin indices assume 1, 2, 3 values.

of tachions, i.e., of hypothetical particles moving faster than light. Since, however, no empirical results confirm such a hypothesis, in the rest of this book it will be systematically ignored.[5]

In the theory of special relativity, geodesic curves (with respect to the connection defined by the Lorentz metric, see chapt. 4, sec. 6) are also identified as histories of freely falling particles of a non–zero rest–mass. This can be thought of as a version of the principle of inertia postulated by special relativity.

The above interpretation of the cone structure of space–time, from the physical point of view, corresponds to its *causal structure*: it informs us on between which domains of space–time information flow (with the help of light signals or other signals slower than light) can take place (see Hawking and Ellis 1973, chapt. 7; Beem and Ehrich 1982, chapt. 2; the causal structure of the Minkowski space–time is presented in the first of these books on pp. 118–124). In this sense, the causal structure of space–time, as considered from a global point of view, has clear cosmological significance.

With the help of the Minkowski metric the distance between two events in space–time can be defined in a natural way. However, let us note that this distance for two different events can be zero, provided that these two events can be joined by a light signal (see below, sec. 5, A). The existence of the Lorentz metric makes it possible to perform space and time measurements in different reference frames. This, in turn, allows one to construct a relativistic mechanics and, in further perspective, also a relativistic electrodynamics.[6]

Everyday laboratory practice of high energy physicists gives a powerful empirical verification of special relativity.

4. The Existence of the Lorentz Structure

4.1. Theorem. A differential manifold $\mathcal{M}$ with an atlas $\{(\mathcal{U}_\alpha, \phi_\alpha)\}$ (at least of the class $\mathcal{C}^1$) admits a (proper) Riemann structure if an only if $\mathcal{M}$ is paracompact.[7] ■

Proof. (1) (*proper*) *Riemann structure* $\Rightarrow$ $\mathcal{M}$ *is paracompact* (directly from the Stone theorem, see, for instance, Engelking 1975, pp. 336 and 342–343; even a stronger statement is true: every metrizable space is paracompact).

(2) $\mathcal{M}$ *is paracompact* $\Rightarrow$ (*proper Riemann structure*). First, let us recall an important notion. Let X be a $\mathcal{C}^r$ manifold. A family of $\mathcal{C}^r$ functions Θ_k, defined on X, is said to be a $\mathcal{C}^r$ *partition of unity on* X if (a) the supports of functions Θ_k,

[5] In older literature, the term "world line" is used to denote the history of a particle or of an observer; this term was coined by Minkowski himself. For a stricter definition of "history" see chapt. 6, sec. 2.

[6] Let us note that the assumption of a finite velocity of light is contained in the causal structure. The numerical value of this velocity has to be taken from experimental results.

[7] Reminder: a topological Hausdorff space X is *paracompact* if into each its open covering $\{U_i\}$ a locally finite covering $\{V_i\}$ can be *inscribed*. A covering $\{V_i\}$ of a topological space X is *locally finite* if for every $x \in X$ there is a neighbourhood having a non–empty intersection only with a finite number of sets $V_i \in \{V_i\}$. Every compact set is paracompact but not vice versa.

supp Θ_k, are compact, (b) for every $x \in X$, there is a finite number of functions Θ_k such that $\Theta_k \neq 0$, (c) $\sum_k \Theta_k(x) = 1$ for each $x \in X$. Further, let $\{(\mathcal{V}_\alpha, \psi_\alpha)\}$ be an atlas on X. A partition of unity on X is said to be *subordinated* to the atlas $\{(V_\alpha, \psi_\alpha\}$ if, for every k, there exists $\mathcal{V}_\alpha$ such that supp $\Theta_k \in \mathcal{V}_\alpha$. Now, let us come back to the proof. Let Θ_i be a partition of unity subordinated to the atlas $\{(\mathcal{U}, \phi_\alpha)\}$ on $\mathcal{M}$. Further, let g_i be a tensor of the type $(0,2)$ on $\mathcal{U}_\alpha$ having, in a certain basis (called a *natural basis*), the coordinates $(g_i)_{kl} = \delta_{kl}$. As it can be easily seen, a tensor $\sum_i \Theta_i g_i$ of the type $(0,2)$ defines a (proper) Riemann structure on $\mathcal{M}$.[8] □

Let us note that the argument contained in the above proof does not hold for a Lorentz structure. The sum $\sum_i \Theta_i g_i$, where g_i are quadratic forms with the index 1, need not be quadratic forms with the index 1. Therefore, one must look for different criteria as far as the existence of a Lorentz structure on $\mathcal{M}$ is concerned. Before proving the theorem giving such a criterion, we must introduce a new concept.

A function $\mathcal{D}$, coordinating to every point x of a differential manifold X an h–dimensional subspace $\mathcal{D}(x)$ of the tangent space $\mathcal{T}_x(\mathcal{M})$ is called an *h–dimensional distribution on X*.[9] If, for every point $x \in X$, there exists a neighbourhood $\mathcal{U}$ of x and $\mathcal{C}^r$ vector fields $\mathcal{V}_1, \ldots, \mathcal{V}_h$ defined on $\mathcal{U}$ such that, for every $p \in \mathcal{U}$, the vectors $\mathcal{V}_1(p), \ldots, \mathcal{V}_h(p)$ form a basis of the subspace $\mathcal{D}(x)$, the distribution $\mathcal{D}(x)$ is said to be of the class $\mathcal{C}^r$. One–dimensional distribution on X is called a *direction field* on X. The concept of h–dimensional distribution turns out to be an efficient tool in the theory of partial differential equations of the first order (see Bishop and Goldberg 1980, chapt. 3. 11–12).

4.2. Theorem. A Lorentz structure exists on a paracompact differential manifold X (at least of the class $\mathcal{C}^1$), if and only if a continuous direction field exists on X. ■

Proof. (1) *Continuous direction field* ⇒ *Lorentz structure.* X is paracompact, and consequently (on the strength of the preceding theorem) a (proper) Riemann structure g exists on X. Let u be a unit vector belonging to the direction field, i.e. $u \in \mathcal{D}(x) \subset \mathcal{T}_x(X)$, where $\mathcal{D}(x)$ is a continuous one–dimensional distribution on X. Let us define

$$\mathcal{G}_x(v,v) = g_x(v,v) - 2(g_x(v,u))^2$$

with $v \in \mathcal{T}_x(X)$. It can be easily checked that it is a Lorentz metric (e.g. in the four–dimensional case, by writing down the above expression in coordinates and substituing: $u^0 = u_0 = 1$, $g_{\alpha\beta} = \delta_{\alpha\beta}$)

(2) *Lorentz structure* ⇒ *continuous direction field.* Let g — as above — be a (proper) Riemann mrtric on X. One of the eingenvectors, to be found by solving the

[8] From the definition of paracompactness it follows that a paracompact manifold can be covered by a denumerable set of local charts. In fact, the implication in the opposite direction is also true (see Geroch 1971, Appendix).

[9] This concept has nothing in common with distribution as generalized functions (in the sense of Schwartz).

eigenvalue problem

$$\mathcal{G}_{\alpha\beta}u^{\beta} = \lambda g_{\alpha\beta}u^{\beta},$$

is timelike; it defines a continuous direction field on X. (See Choquet–Bruhat et al. 1982, pp. 292–293; Geroch 1971). □

Two comments should be given. Strictly speaking, in order to have a $\mathcal{C}^r$ Lorentz structure on X one should have a $\mathcal{C}^r$ direction field on X: the existence of a *continuous* direction field on X implies only the existence of a *continuous* Lorentz structure on X. However, this is not a serious limitation since, as shown by Steenrod (1951, p. 25 seq.), a continuous direction field on X can be approximated by a smooth direction field (in fact similar approximations are valid as far as continuous cross–sections of any vector bundles are concerned).

Second, the existence of a continuous direction field on X is not only a necessary condition for the existence of a Lorentz structure on X, but also a Lorentz structure on X can always be chosen in such a way that the direction field in question be timelike in this structure.

The existence of a direction field on a manifold is its topological property. As it is known from topology, every non–compact (but paracompact) manifold admits the existence of a continuous direction field; however, a compact manifold admits the existence of a continuous direction field if and only if its Euler–Poincaré characteristic vanishes. It turns out that — on the strength of one of a few possible Euler–Poincaré characteristic definitions — the existence of a Lorentz structure on compact manifolds is strictly connected with the structure of differential forms on these manifolds. Here is an outline of the idea.

Let the space of all p–forms (of the class $\mathcal{C}^k$) on a manifold X be denoted by $\Lambda^p(X)$ (see chapter 1, sec. 4). Since p–forms are antisymmetric, one must have $p \leq n$, where $n = \dim(X)$. The space $\Lambda^p(X)$ is closed with respect to the addition of p–forms and their multiplication by functions (of $\mathcal{C}^k$ class). As it is well–known, an *external product* can be defined

$$\wedge\colon (\Lambda^p(X), \Lambda^q(X)) \longrightarrow \Lambda^{p+q}(X),$$

allowing one to multiply a p–form α by a q–form β to obtain a $(p+q)$–form $\alpha \wedge \beta$ (more about differential forms on a manifold, see: Flanders 1963; Choquet–Bruhat et al. 1982, chapter 4, Bishop and Goldberg 1980, chapt. 4).

The set $\Lambda(X)$ of all forms of any order on X, together with an external product defined in it, is called an *external algebra* or a *Grassman algebra*. A Grassman algebra is a graded algebra, i.e. a set of sets $\{\Lambda^p(X)\}$, indexed by the integers $p = 0, 1, \ldots, n$, together with an associated bilinear mapping $\Lambda(X) \times \Lambda(X) \to \Lambda(X)$; in our case — an external product. An external differentation

$$d\colon \Lambda^p(X) \longrightarrow \Lambda^{p+1}(X)$$

with the property $dd = d^2 = 0$, allows one to treat $\Lambda(X)$ as a *differential* graded algebra. Let us note that the sets $\Lambda^p(X)$ are vector spaces (over $\mathbf{R}$).

Let $\omega \in \Lambda^p(X)$. ω is said to be a *closed form* or a *co-cycle*, if $d\omega = 0$. ω is called an *exact form* or a *co-boundary*, if $\omega = d\Theta$ (of course $\Theta \in \Lambda^{p-1}(X)$).

Evidently, every exact form is closed (since $d^2 = 0$), but not vice versa. The study of closed forms that are not exact brings important information about topology of a given space in the following way.

Let the vector space of closed p–forms on a manifold be denoted by $\mathcal{Z}^p$, and the vector space of exact p–forms on X by $\mathcal{B}^p$. Evidently, one has $\mathcal{B}^p \subset \mathcal{Z}^p$. The quotient space $\mathcal{H}^p = \mathcal{Z}^p/\mathcal{B}^p$ is called a *p-cohomology space* of the manifold X. Since $\mathcal{H}^p$ is an Abelian group with respect to addition it is also called a *cohomology group of p^{th} order* of X. Equivalence classes of closed p–forms are elements of $\mathcal{H}^p$: two closed forms ω_1 and ω_2 are considered to be equivalent if they differ by an exact form, i.e., if $\omega_1 - \omega_2 = d\Theta$.

The dimension of the p^{th} order cohomology group of X is called a p^{th} *Betti number* of the manifold X and is denoted by b^p. The sum $\chi(X) = \sum_{q=0}^{n}(-1)^q b^q$, where $n = \dim(X)$ is called an *Euler–Poincaré characteristic of this manifold*.[10] A compact differential manifold X can be equipped with a Lorentz structure if and only if $\chi(X) = 0$.

Let us note that the vanishing of the Euler–Poincaré characteristic of a compact manifold X is a necessary and sufficient condition of the *global* existence of a Lorentz structure on it. A direction field, and consequently a Lorentz structure, *locally* always exist on X.

5. Comments and Remarks

A. As is well–known, a (proper) Riemann metric on a differential manifold X defines, in a natural way, a (*Riemann*) *distance function* $d_R: X \times X \to [0, \infty)$ on X, satisfying the usual axioms (including the *triangle inequality*). The topology determined by this function, the so–called *metric topology*, is identical with the original manifold topology (see chapt. 1, sec. 2). No Lorentz metric has these nice properties. In terms of the Lorentz structure, one can define a (*Lorentz*) *distance function* $d_L: X \times X \to [0, \infty]$ (which can assume infinite values!), but it satisfies axioms different from those satisfied by d_R (in particular, it satisfies the so–called *reversed triangle inequality*). Any distance measured by d_L along non–timelike curves is, from definition, equal to zero. A topology determined by d_L is essentially weaker than the manifold topology on X. Unfortunately, entering into details would require new concepts going beyond the scope of the present book; the interested reader should consult the book by Beem and Ehrlich (1981, chapters 1 and 3).

[10] The Betti numbers and the Euler–Poincaré characteristic of a given manifold can be also defined dually in terms of a triangulation of this manifold. Dual with respect to closed and exact forms are the so–called *cycles* and *boundaries*, respectively. In a similar manner one defines a homology group (or space) $\mathcal{H}_p$ as a quotient space $\mathcal{Z}_p/\mathcal{B}_p$, where $\mathcal{Z}_p$ is a vector space of cycles and $\mathcal{B}_p$ a vector space of boundaries (of the dimension p).

B. As we have seen, the existence of the Lorentz structure on a (paracompact) manifold is equivalent to the existence of a continuous direction field; moreover, the Lorentz structure can always be chosen so as to make the direction field timelike in this structure. This fact seems to have a deep cosmological significance; one can express it by saying that the existence of a Lorentz structure on a manifold is equivalent to the existence of a *local arrowless* time on it. The existence of a continuous timelike direction field on a manifold could be interpreted as a geometric counterpart of the fact that each point of the manifold feels a lapse of a local time. This lapse, however, is determined by a direction and not by a vector, i.e. two possible arrows along the same direction (the past and the future) are equivalent (the present structure does not distinguish between them). It is truly an arrowless time. And it is also a local time: a timelike direction field, besides the rather tolerant requirement of continuity, does not impose any global conditions; in particular, it provides no conceptual tools which could serve to compare "local times" at different points of the space–time manifold.

To sum up: a paracompact differential manifold can carry a Lorentz structure if any only if a continuous direction field exists on it, and this direction field, after the imposing a Lorentz structure on the manifold, can be interpreted as a local arrowless time (see Heller 1986, chapt. 4).

C. If one agrees that without a Lorentz structure physics cannot be done on a manifold (see introduction to the present section), one must admit that the existence of at least a local arrowless time is a condition for the *possibility of physics*. A similar idea was expressed by C. von Weizsäcker (1972). In his opinion, science establishes laws which explain experiments performed in the past, but are verified (or falsified) by predicting outcomes of future experiments, and by comparing them with the results of experiments when they are already present results; in this sense time is a precondition of physics. If one remembers that, from the methodological point of view, retrodictions are as important for science as are predictions (although psychologically predictions play doubtlessly a more important role), one is entitled to assert that the existence of an (at least) arrowless time is a precondition for physics (see Heller 1986, chapt. 4).

D. Although a compact differential manifold with the vanishing Euler–Poincaré characteristic admits the existence of a Lorentz structure, it leads to some pathologies when considered as a model of space–time. Bass and Witten (1957) have shown that a four–dimensional compact manifold with the vanishing Euler–Poincaré characteristic cannot be simply connected.[11] Moreover, it turns out that in every such manifold at least one *closed* timelike curve must exist (see Beem and Ehrlich 1981, p. 23). Physically, this would mean the existence of a closed history of an observer or of a non–zero rest–mass particle (i.e., of a history which would endlessly repeat itself again and again). It can hardly be seen how one could reasonably make this consistent with requirements of the macroscopic causality (for interesting remarks

[11] Its first Betti number must be different from zero.

concerning this problem see Hawking 1971, p. 396; Hawking and Ellis 1973, p. 189). No four–dimensional compact manifold equipped with a Lorentz structure can be simply connected. However, such a manifold could be treated as a non–compact Lorentz manifold in which some points have been identified. More precisely, for any n–dimensional manifold X_0, there is a unique simply connected manifold X covering the manifold X_0; X is called a *universal covering* of X_0 or a *manifold universally covering* the manifold X_0 (for definition see Spanier 1966, chapt. 2; Rychtmyer 1981, chapt. 24). Therefore, from the point of view of local properties, instead of a four–dimensional compact manifold one could take its universal covering as a model of the physical space–time. This covering is non–compact and simply connected (see Hawking and Ellis 1973, p. 190).

E. Now, we are ready to take up the question posed in chapter 1 (sec. 7, A) concerning the mutual relationship between a coordinate system and a reference frame. Coordinate system is a mathematical concept which is equivalent to that of a local chart belonging to an atlas that defines a differential manifold structure. A reference frame is a physical counterpart of a coordinate system; however, not every transformation of coordinates would be treated by a physicist as a transition to a new coordinate system. For instance, an affine transformation of the form $x \mapsto ax + b$, $a, b \in \mathbf{R}$ (changing to new units and new zero point on a scale) or the transition from Cartesian to polar coordinates would not be considered as going to a new reference frame. On the other hand, "rotating coordinates" with respect to previous ones would be regarded as being connected with a new frame of reference. Clearly, the difference between these two classes of coordinate transformations consists in the fact that transformations of the former class are independent of time whereas the transformation of the latter class do depend on time. Therefore, the definition of a reference frame must contain: (a) a definition of rest (with respect to this reference frame) and simultaneity; (b) a definition of the metric structure of time (t) together with a unit of time, and (c) a definition of the metric structure of instantaneous spaces ($t = \text{const}$) together with a distance unit. Owing to these definitions one can meaningfully speak separately of time and separately of space, but only with respect to a given reference frame (see Dixon 1978, chapt. 1, sec. 2).

F. Clarke (1970) has shown that every Lorentz manifold can be smoothly embedded into a flat (pseudo)Euclidean space of suitably many dimensions; this embedding can be of an arbitrarily high differentiability class. In particular, every compact four–dimensional Lorentz manifold can be smoothly embedded into the 48–dimensional pseudo–Euclidean space with the index 2 (2 minuses and 46 pluses), and every non–compact four–dimensional Lorentz manifold can be smoothly embedded into the 89 dimensional Euclidean space with the index 2 (2 minuses and 87 pluses).

Bibliographical Note. From among books on differential geometry, enumerated in the *Bibliographical Note* after chapter 1, only Choquet–Bruhat et al. (1982), and Bishop and Goldberg (1980) deal extensively with Lorentz manifolds. The monographs by Hawking and Ellis (1973) and by Beem and Ehrlich (1981) are entirely devoted to Lorentz manifolds and their applications to

space–time geometry (see also Heller 1981b, 1986 chapt. 3 and 4). The book by O'Neill (1983) should be especially recommended; it develops, step by step, the geometry of spaces with Lorentz metrics, and discusses they use in both special and general relativity. This book can also be treated as a kind of introduction to more advanced texts (like, for example, Hawking and Ellis (1973) or Beem and Ehrlich (1981)).

The reader is urged to make a closer acquaintance with the theory of special relativity. Its basic ideas can be found in every modern textbook of physics. A good introduction, with a penetrating discussion of the conceptual framework of this theory, is Taylor and Wheeler (1966). First chapters of Schutz (1985) are equally good but much shorter. There exist books on special relativity of the monographic character, for example: Aharoni (1959), Synge (1965), Dixon (1978). The history of special relativity can be found in Miller (1981); this book contains also a new English translation of Einstein's original paper (1905) and its extensive analysis.

Chapter 3

Method of Fibre Bundles

0. Introduction

Fibre bundles functioned in physics for quite a time before their strict definition was formulated by mathematicians. For instance, as it is well–known, the Lagrangian of a dynamical system is a real function of coordinates and velocities of particles constituting this system. In other words, if $\mathcal{M}$ is an n–dimensional configuration space of a given system, the Lagrangian is defined on the space of all coordinates and (generalized) velocities, that is, on all tangent vectors to $\mathcal{M}$ at all points x of $\mathcal{M}$. Such a space is called a *tangent bundle* over $\mathcal{M}$, denoted by $\mathcal{T}(\mathcal{M})$, and it evidently has $2n$ dimensions. A tangent bundle $\mathcal{T}(\mathcal{M})$ is in fact a union of tangent spaces $\mathcal{T}_x(\mathcal{M})$ to the manifold $\mathcal{M}$ at all its points $x \in \mathcal{M}$. The space $\mathcal{T}(\mathcal{M})$ can be given the structure of a differential manifold; $\mathcal{T}(\mathcal{M})$ equipped with this structure is called a *bundle space*, and $\mathcal{M}$ its *base space*. A dynamical system can also be investigated with the help of the Hamilton method. The Hamiltonian of the system is a function of coordinates and (generalized) momenta. Since momenta are linear combinations of (generalized) velocities, one says that the Hamiltonian is defined on a *cotangent bundle* over the configuration space $\mathcal{M}$, where the cotangent bundle over $\mathcal{M}$, denoted by $\mathcal{T}_x^*(\mathcal{M})$, is to be understood as a union of dual spaces $\mathcal{T}_x^*(\mathcal{M})$ to the spaces $\mathcal{T}_x(\mathcal{M})$ for all $x \in \mathcal{M}$. The space $\mathcal{T}^*(\mathcal{M})$ can also be equipped with the structure of a differential manifold. (For classical mechanics in the language of bundles, see Abraham and Marsden 1979, Thirring 1978, Arnold 1974.)

The Kaluza–Klein theory can serve as another, perhaps even more sophisticated, example of the functioning of fibre bundles in physics before they were discovered by mathematicians. Kaluza and Klein, yet in 1921, aimed at unifying gravitational and electromagnetic interactions within the framework of a five–dimensional Riemann space on which a one–parameter isometry group was acting (see, for instance, Bergman 1976, chapters 17 and 18). *Ex post* it has turned out that the space, postulated by Kaluza and Klein, is the so–called *circle bundle*, and the electromagnetic potentials, in their approach, are connection forms in it (see, Trautman 1980).

After the Second World War the theory of fibre bundless began gradually to dominate differential geometry and many chapters of topology as well. Today, it constitutes a standard method of the so–called global analysis. After passing through the hands of mathematicians and obtaining a precise, fully formalized form, the fibre bundle method has come back to physics. First, as an efficient tool not only

introducing an order into individual physical theories (such as classical mechanics, see above), but also showing how some physical theories can be arranged into evolutionary chains revealing a certain logic of their development (see Appendix). Later on, the theory of fibre bundles was recognized as a suitable mathematical apparatus allowing one to geometrize gauge theories (see, e.g. Drechsler and Mayer 1977, Nash and Sen 1983). It has turned out that the case of the Kaluza–Klein theory was not an exception but rather a signal of a more general rule: gauge potentials, so–called by physicists, are essentially connection forms in principal fibre bundles. If one takes into account the fact that the unification of the electromagnetic and lepton (weak) interactions, achieved by Weinberg and Salam, has been obtained within the framework of the gauge scheme, and that there are strong reasons to believe that the same scheme will work in unifying strong (hadron) interaction with the electroweak one (the so called grand unification), and perhaps also with gravity (superunification), one can hardly overestimate the role of fibre bundles in contemporary physics (see, for instance, Taylor 1978, Ross 1981).

Today it is also known that grand unification and, even more so, superunification can work only in an environment of extremely high energies. One estimates that the energies needed for grand unification and superunification are of order of 10^{14} GeV and 10^{19} GeV respectively. Such a high energy can be supplied only by the Universe during its very early evolution. According to the standard cosmological model the energy necessary for grand unification was obtainable in the Universe 10^{-33} s after the initial singularity, and the energy necessary for superunification 10^{-44} s after the initial singularity. In this way, the unification of physics seems to be inseparably connected with cosmology.

Einstein's hope for unifying physics in a geometric framework was not in vain. It has only turned out that the traditional space time geometry is too poor for this end. Physical events cannot be mere points devoid of any structure. The heart of the matter consists in that that such a structure can be given to them by considering the set of suitable geometric objects (tangent spaces, all possible frames, an so on) at a given point of space–time, i.e. by considering a suitable fibre over this space time point. Collecting together all such fibres (over all points of space time) one obtains a corresponding fibre bundle with space–time as its base manifold.

The reasons presented above sufficiently justify our aim to present general relativity, with the perspective of its future cosmological applications, from the very beginning in terms of fibre bundles. We shall do that in the next chapter; in the present chapter we focus on the fibre bundle method itself. Section 1 introduces the concept of fibre bundles, and sec. 2 specifies the principal and associated fibre bundles. In sec. 3, the reader will find an outline of the theory of connections in fibre bundles; because of its importance in physics this theory is presented in a greater detail. In sec. 4, the geometry of a principal fibre bundle (connection and its curvature, structure equations, etc.) is constructed, and in sec. 5 this geometry is "projected down" to the base manifold of the bundle. The usual *Comments and Remarks* are

replaced in this chapter by a concise presentation of some applications of the fibre bundle method in gauge theories in physics.

1. Fibre Bundles

The fibre bundle concept is a generalisation of the Cartesian product. For instance, a cylinder can be presented as $\mathcal{S}^1 \times \mathcal{I}$, where $\mathcal{S}^1$ is a one–sphere and $\mathcal{I} \subset \mathbf{R}^1$, but something like that cannot be done with respect to a Möbius strip. However, it is not difficult to see that Möbius strip is "locally" a Cartesian product: if U is a proper subset of $\mathcal{S}^1$, then $U \times \mathcal{I}$ is "a part" of a Möbius strip. We say that Möbius strip is a fibre bundle over $\mathcal{S}^1$ with $\mathcal{I}$ as a typical fibre. This situation can be generalised in the following chain of definitions.

1.1. Definition. The triple $(\mathcal{E}, \mathcal{B}, \pi)$, where $\mathcal{E}$ and $\mathcal{B}$ are topological spaces, and $\pi: \mathcal{E} \to \mathcal{B}$ a continuous surjective mapping called *projection*, is said to be a *bundle*.

1.2. Example. Let $\mathcal{A}$ and $\mathcal{B}$ be topological spaces. The projection $\pi_A: \mathcal{A} \times \mathcal{B} \to \mathcal{A}$ is defined by $\pi_A(x_A, x_B) = x_A$, $x_A \in \mathcal{A}$, $x \in \mathcal{B}$. $(\mathcal{A} \times \mathcal{B}, \mathcal{A}, \pi_A)$ is a bundle.

A bundle which is a Cartesian product (like that of example 1.2) is called a *trivial* (or *product*) *bundle*.

1.3. Definition. A *fibre bundle* is a set of five elements $\mathcal{E}(\mathcal{B}) = (\mathcal{E}, \mathcal{B}, \pi, \mathcal{G}, \mathcal{F})$, where $(\mathcal{E}, \mathcal{B}, \pi)$ is a bundle, $\mathcal{E}$ is called a *bundle space* (or a *total space*), $\mathcal{B}$ is a *base* (or a *base space*). It is assumed that $\mathcal{B}$ has an open covering $\{U_j; j \in \mathcal{J}\}$. The set $\pi^{-1}(x) \equiv \mathcal{F}_x$, $x \in \mathcal{B}$, is called a *fibre over* x. It is assumed that fibres $\mathcal{F}_x$, for every $x \in \mathcal{B}$, are homeomorphic to the topological space $\mathcal{F}$, called a *typical fibre*. $\mathcal{G}$ is a group acting on $\mathcal{F}$ to the left; it is called a *structural group* of the fibre bundle. The following conditions are supposed to be satisfied:

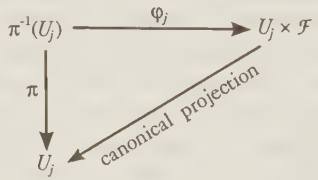
(1) The fibre bundle is *locally trivial*, i.e., there exists a homeomorphism

$$\varphi_j: \ \pi^{-1}(U_j) \longrightarrow U_j \times \mathcal{F}$$

given by $\varphi_j(p) = (\pi(p), \phi_j(p))$, $p \in \pi^{-1}(U_j)$, where

$$\phi_j\big|_{\mathcal{F}_x} \equiv \phi_{j,x}: \ \mathcal{F}_x \longrightarrow \mathcal{F}$$

is a surjective homeomorphism, and the diagram



is commutative.

(2) The *consistency condition* holds, i.e., for every $x \in U_j \cap U_k$ and every $j, k \in \mathcal{J}$, the homeomorphic mapping

$$\phi_{k,x} \circ \phi_{j,k}^{-1}\colon \mathcal{F} \longrightarrow \mathcal{F}$$

is an element of the structural group $\mathcal{G}$ (see fig. 3.1).

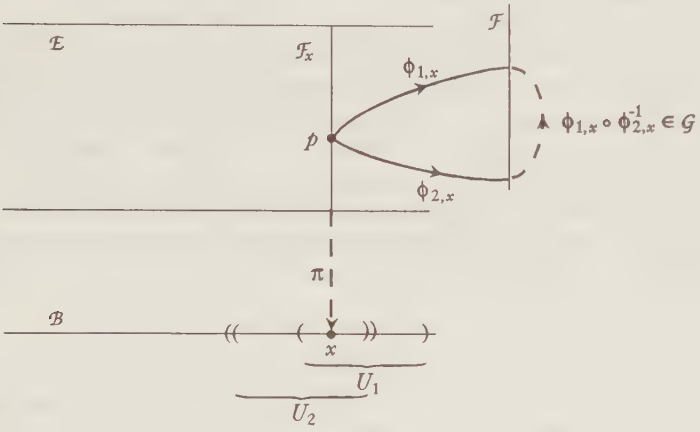


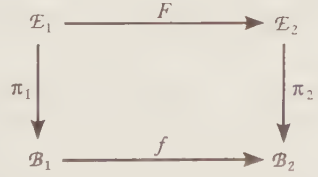
Fig. 3.1. Compatibility condition.

(3) The mappings

$$\gamma_{jk}\colon U_j \cap U_k \longrightarrow \mathcal{G},$$

defined by $\gamma_{jk}(x) = \phi_{j,x} \circ \phi_{k,x}^{-1}$, $x \in U_j \cap U_k$, are continuous; γ_{ij} are called *transition functions*.

1.4. Definition. Two bundles $(\mathcal{E}_i, \mathcal{B}_i, \pi_i)$, $i = 1, 2$, are said to be *equivalent* if there exist mappings $F\colon \mathcal{E}_1 \to \mathcal{E}_2$ and $f\colon \mathcal{B}_1 \to \mathcal{B}_2$ that preserve the bundle structure, i.e., such that the diagram



is commutative. The pair (F, f) is said to be a *bundle morphism*. The equivalence of bundles plays a similar role in the fibre bundle theory as the equivalence of atlases does

in the theory of manifolds[1] (see chapt. 1, sec. 2): two fibre bundles $(\mathcal{E}_i, \mathcal{B}_i, \pi_i, \mathcal{G}_i, \mathcal{F}_i)$, $i = 1, 2$, are considered to be equivalent if treated locally, as the bundles $(U_i \times \mathcal{F}_i, U_i, \pi_i)$ are equivalent; additionally one assumes that U_i belong to the open covering $\{U_j;\ j \in \mathcal{J}\}$ of $\mathcal{B}_i$ and $\{U_i\}$ is consistent with the fibre bundle structure, i.e., it satisfies conditions (1) and (2) of the definition of fibre bundle.

The sets $\{U_j, \varphi_j\}$ appearing in the definition of a fibre bundle (condition (1)) are called *local trivializations* of the bundle. If the structural group $\mathcal{G}$ has only one element, the bundle is called *trivializable*.

1.5. Definition. A fibre bundle $\mathcal{E}(\mathcal{B}) = (\mathcal{E}, \mathcal{B}, \pi, \mathcal{G}, \mathcal{F})$ is said to be a *differential fibre bundle (of the class $\mathcal{C}^k$)* if $\mathcal{E}$, $\mathcal{B}$ and $\mathcal{F}$ are differential manifolds (of the class $\mathcal{C}^k$), the sets U_j covering $\mathcal{B}$ coincide with domains of local charts on $\mathcal{B}$ as a manifold, $\mathcal{G}$ is a Lie group, and mappings π and γ_{ik} are differentiable (of the class $\mathcal{C}^k$). In the following, without saying this directly, we shall consider fibre bundles of the class $\mathcal{C}^\infty$.

1.6. Example. Let $\mathcal{S}^2$ be a two–sphere with a chosen orientation. A pair of unit orthogonal vectors $(e_1, e_2)_x$, tangent to $\mathcal{S}^2$ at $x \in \mathcal{S}^2$, is called a *diad*. The set of all diads having orientation consistent with that of $\mathcal{S}^2$ is denoted by $\mathcal{P}$. Let us define the projection $\pi\colon \mathcal{P} \to \mathcal{S}^2$ by $\pi(e_1, e_2)_x = x$. If $(e_1, e_2)_x$ is a diad at a point $x \in \mathcal{S}^2$, then the transformation

$$e_1' = e_1 \cos\varphi + e_2 \sin\varphi,$$
$$e_2' = -e_1 \sin\varphi + e_2 \cos\varphi$$

gives another diad at the same point. Such transformations are elements of the group $\mathcal{SO}(2)$ which turns out to be the structural group of the bundle $\mathcal{P}(\mathcal{S}^2) = (\mathcal{P}, \mathcal{S}^2, \pi, \mathcal{SO}(2))$. The typical fibre $\mathcal{F}$ is identical with the group $\mathcal{SO}(2)$ treated as a manifold (see below, sec. 2). On the strength of the Brouwer theorem about the impossibility to smoothly comb a vector field on a sphere of an even dimension, the fibre bundle $\mathcal{P}(\mathcal{S}^2)$ is not trivial. (See Trautman 1978, pp. 6–7.)

1.7. Example. Möbius strip (see Choquet–Bruhat et al. 1982, pp. 126–127.)

If a fibre bundle $\mathcal{E}(\mathcal{B})$ with a structural group $\mathcal{G}$, after replacing $\mathcal{G}$ by its subgroup $\mathcal{G}_1$, leads to the equivalent bundle structure, the structural group $\mathcal{G}$ is said to be *reducible* to the subgroup $\mathcal{G}$, or the bundle $\mathcal{E}(\mathcal{B})$ with the structural group $\mathcal{G}_1$ is said to be a *reduction* of the bundle $\mathcal{E}(\mathcal{B})$ with the structural group $\mathcal{G}$. If such a reduction occurs, then there exists a family of local trivializatons such that the transition functions γ_{ij} corresponding to them assume values in $\mathcal{G}_1$ (see chapt. 4, sec. 5).

[1] Bundles together with bundle morphisms define the *category of bundles*. By suitably generalizing the morphism concept (by taking into account the structure induced by the existence of a typical fibre and of a structural group) one can define the *category of fibre bundles* (see Sulanke and Wintgen 1972).

2. Principal and Associated Fibre Bundles

2.1. Definition. A fibre bundle $\mathcal{E}(\mathcal{B}) = (\mathcal{E}, \mathcal{B}, \pi, \mathcal{G}, \mathcal{F})$, the structural group of which, $\mathcal{G}$, acts to the left on the typical fibre $\mathcal{F}$, i.e., $L_g(f) = gf$, $g \in \mathcal{G}$, $f \in \mathcal{F}$, and such that $\mathcal{F}$ is identical with $\mathcal{G}$ (understood as a manifold), is called a *principal fibre bundle*.

2.2. Example. Let ρ_x be a basis in $\mathcal{T}_x(X^n)$, i.e., $\rho_x = (e_i)$, $i = 1, \ldots, n$, where e_i are linearly independent vectors; these vectors, in turn, can be expressed as linear combinations of vectors $\tilde{e}_i$ which form another basis in $\mathcal{T}_x(X)$. Of course, one has

$$e_i = a^j_i \tilde{e}_j, \quad (a^j_i) = a \in \mathcal{GL}(\mathbf{R}^n).$$

Thus, there is a bijection between the set of all bases in $\mathcal{T}_x(X^n)$ and the group $\mathcal{GL}(\mathbf{R}^n)$. The group $\mathcal{GL}(\mathbf{R}^n)$ is a Lie group and, as a manifold, it can be identified with an open subset in $\mathbf{R}^{n^2}$. Let us define $\mathcal{F}(X^n)$ to be the set of pairs (x, ρ_x), for every $x \in X^n$, and the projection $\pi(x, \rho_x) = x$. In this way, we have obtained the fibre bundle $\mathcal{F}(X^n) = (\mathcal{F}(X^n), X^n, \pi, \mathcal{GL}(\mathbf{R}^n))$ which is called the *fibre bundle of linear frames*.

2.3. Example. The fibre bundle of diads over a sphere is a principal fibre bundle (see example 1.6).

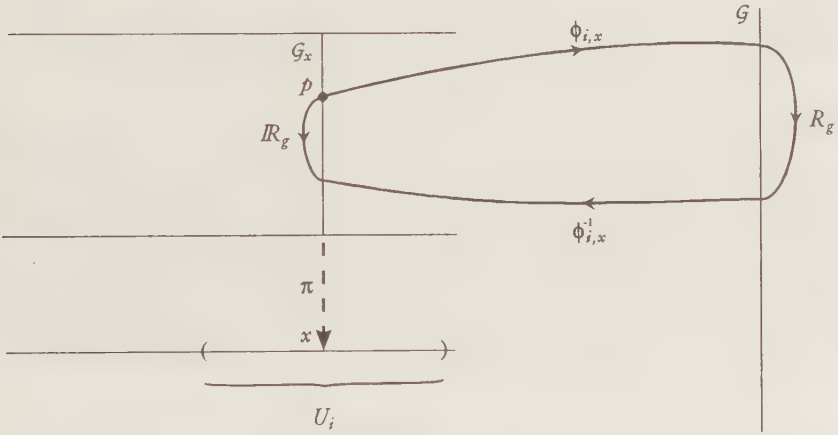


Fig. 3.2. Right action of the structural group $\mathcal{G}$ on a fibre $\mathcal{G}_x$ of a principal fibre bundle.

Let $\mathcal{E}(X) = (\mathcal{E}, X, \pi, \mathcal{G})$ be a principal fibre bundle. Let further $p \in \mathcal{G}_x \equiv \pi^{-1}(x)$, $x \in U_i$, with U_i belonging to an open covering of X which defines the fibre bundle structure. Let us remember that $\phi_{i,x}: \mathcal{G}_x \to \mathcal{G}$, and let $g_i \equiv \phi_{i,x}(p)$. We define the *right action* R_g, $g \in \mathcal{G}$, of the structural group $\mathcal{G}$ on the fibre bundle in the following way (see fig. 3.2)

$$(R_g p)_i = \phi^{-1}_{i,x} \circ R_g \circ \phi_{i,x} = \phi^{-1}_{i,x}(R_g g_i) = \phi^{-1}_{i,x}(g_i g).$$

The definition is consistent with the fibre bundle structure, i.e., for $p \in \pi^{-1}(U_i \cap U_j)$, one has $(\mathbb{R}_g p)_i = (\mathbb{R}_g p)_j$ (the reader is urged to prove this as an exercise, see Choquet-Bruhat et al. 1982, p. 130). One can easily see that $\mathbb{R}_g$ acts (transitively) on a single fibre, i.e., $(p \in \mathcal{G}_x) \Rightarrow (\mathbb{R}_g p \in \mathcal{G}_x)$, $x = \pi(p)$.

2.4. Definition. A fibre bundle such that its typical fibre is a vector space and its structural group is a linear group is said to be a *vector fibre bundle*. A vector fibre bundle $\mathcal{E}(X) = (\mathcal{E}, X, \pi, \mathcal{F}, \mathcal{G})$ is said to be *associated* with the principal fibre bundle $\mathcal{P}(X) = (\mathcal{P}, X, \pi, \mathcal{G})$ if there is a representation[2] $\rho: \mathcal{G} \to \mathbb{L}(\mathcal{F}, \mathcal{F})$ of the structural group $\mathcal{G}$ in the space $\mathcal{F}$ such that the transition functions $\gamma^1_{i,j}$ of the associated bundle $\mathcal{E}(X)$ are images (under ρ) of the corresponding transition functions of the principal bundle $\mathcal{P}(X)$, i.e., $\gamma^1_{ij} = \rho(\gamma_{ij})$, or more precisely: Let $\{U_i, \psi_i\}$ be local trivializations of the principal bundle $\mathcal{P}(X)$, and $\{U_i, \varphi_i\}$ local trivializations of the associated bundle $\mathcal{E}(X)$. In such a case the transition functions γ_{ij} of the principal bundle $\mathcal{P}(X)$ are given by

$$\gamma_{ij}\colon U_i \cap U_j \longrightarrow \mathcal{G},$$
$$x \longmapsto \gamma_{ij}(x) = \Psi_{k,x} \circ \Psi_{i,x} \in \mathcal{G},$$

where $\Psi_{k,x}\colon \mathcal{G}_x \to \mathcal{G}$, and the transition functions γ^1_{ij} of the associated bundle $\mathcal{E}(X)$ are given by

$$\gamma^1_{ij}(x) = \phi_{j,x} \circ \phi^{-1}_{i,x} = \rho(\gamma_{ij}(x)) \in \mathbb{L}(\mathcal{F}, \mathcal{F}),$$

where $\phi_{i,x}\colon \pi^{-1}(x) \to \mathcal{F}$.

Sometimes one stresses directly that a given bundle $\mathcal{E}(X)$ is associated with a principal bundle $\mathcal{P}(X)$ via the representation ρ of the structural group. Having a vector bundle $\mathcal{E}(X)$ it is always possible to construct effectively a fibre bundle $\mathcal{P}(X)$ with which $\mathcal{E}(X)$ is associated (see Choquet-Bruhat et al. 1982, p. 368).

2.5. Example. Let $\mathcal{T}(X^n)$ be the space of pairs (x, v_x), $x \in X^n$, $v_x \in \mathcal{T}(X^n)$, and let the projection be defined as $\pi(x, v_x) = x$. Let us further suppose that $\{(U_j, \psi_j)\}$ is an atlas on the manifold X^n. Let us consider two local charts (U_1, ψ_1) and (U_2, ψ_2), and $x \in U_1 \cap U_2$, $v \in \mathcal{T}(X^n)$. We have

$$\psi'_{1,x}\colon\ v_x \longmapsto V_1 \in \mathbf{R}^n$$
$$\psi'_{2,x}\colon\ v_x \longmapsto V_2 \in \mathbf{R}^n$$

The transformation $V_1 \to V_2$ is given by

$$\psi_{1,x} \circ \psi^{-1}_{2,x}\colon\ \mathbf{R}^n \longrightarrow \mathbf{R}^n \in \mathcal{GL}(\mathbf{R}^n).$$

In this way a fibre bundle $\mathcal{T}(X^n) = (\mathcal{T}(X^n), X^n, \pi_1, \mathcal{GL}(\mathbf{R}^n), \mathbf{R}^n)$ has been defined; it is called a *tangent bundle* to the manifold X^n. Since there is a representation

[2] We remind that a *representation* of a group $\mathcal{G}$ on a space $\mathcal{F}$ is a homeomorphism $\mathcal{G} \to \mathbb{L}(\mathcal{F}, \mathcal{F})$, where $\mathbb{L}(\mathcal{F}, \mathcal{F})$ denotes the set of linear transformations $f\colon \mathcal{F} \to \mathcal{F}$.

$\rho: \mathcal{GL}(\mathbf{R}^n) \to \mathbb{L}(\mathbf{R}^n, \mathbf{R}^n)$ the tangent bundle is associated with the bundle of frames $\mathcal{F}(X^n) = (\mathcal{F}(X^n), X^n, \pi, \mathcal{GL}(\mathbf{R}^n))$.

A $\mathcal{C}^r$ mapping $f: \mathcal{B} \to \mathcal{E}$ such that $\pi \circ f = \mathrm{id}$ is called a $\mathcal{C}^r$ *cross-section* of the bundle $(\mathcal{E}, \mathcal{B}, \pi)$. A $\mathcal{C}^r$ *field of contravariant vectors* on a manifold X^n can be defined as a $\mathcal{C}^r$ cross-section of the tangent bundle $\mathcal{T}(X^n)$.

2.6. Example. Let $\mathcal{T}^*(X^n)$ denote the space of pairs (x, ω_x), $x \in X^n$, $\omega_x \in \mathcal{T}_x^*(X^n)$, and let the projection be defined as $\pi_2(x, \omega_x) = x$. The fibre bundle $\mathcal{T}_x^*(X^n) = (\mathcal{T}_x^*(X^n), X^n, \pi, \mathcal{GL}(\mathbf{R}^n), \mathbf{R}^n)$ is called the *cotangent bundle* to the manifold X^n. It is also associated with the frame bundle. A cross-section of the cotangent bundle is a *field of covariant vectors* or a 1-*form* on X^n.

2.7. Example. By generalising examples 2.5 and 2.6 we define a *tensor fibre bundle* of the type (q, p) over a manifold X^n as $\mathcal{T}^q{}_p(X^n) = (\mathcal{T}^p{}_q(X^n), X^n, \pi, \mathcal{G}, \mathcal{F})$, where

$$\mathcal{T}^p{}_q(X^n) \quad \text{is the space of pairs} \quad (x, t^q{}_p(x)), x \in X^n,$$
$$t^q{}_p(x) \in \otimes^q \mathcal{T}_x(X^n) \otimes^p \mathcal{T}_x^*(X^n),$$
$$\pi(x, t^q{}_p(x)) = x,$$
$$\mathcal{G} = \otimes^{p+q} \mathcal{GL}(\mathbf{R}^n), \quad \mathcal{F} = \otimes^q \mathbf{R}^n \otimes^p \mathbf{R}^{*n}.$$

It is a fibre associated with the frame bundle. A cross-section of a tensor bundle of the type (q, p) is a tensor field of the type (q, p) on the manifold X^n.

Tensor bundles over a paracompact manifold always admit cross-sections (see Choquet-Bruhat et al. 1982, pp. 385–386).

The concept of reduction of a principal fibre bundle, introduced in the preceding section, can be reformulated in the following way. A principal fibre bundle $\mathcal{E}(X) = (\mathcal{E}, X, \pi, \mathcal{G})$ is said to be *reducible* to a principal fibre bundle $\mathcal{E}_1(X) = (\mathcal{E}_1, X, \pi_1, \mathcal{G}_1)$, where $\mathcal{E}_1 \subset \mathcal{E}$, and $\mathcal{G}_1$ is a subgroup of $\mathcal{G}$, if there is an injective mapping $f: \mathcal{E}_1 \to \mathcal{E}$ which is a morphism of the bundles $\mathcal{E}_1(X)$ and $\mathcal{E}(X)$ and which commutes with the action of the group $\mathcal{G}_1$, i.e., $f(\mathbb{R}_g p) = \mathbb{R}_g f(p)$, $g \in \mathcal{G}$, for every $p \in \mathcal{E}_1$ and $\pi f(p) = \pi_1(p)$, for every $p \in \mathcal{E}_1$.

3. Connections in a Principal Fibre Bundle

Connection in a principal fibre bundle is a structure of crucial importance for our further considerations. Therefore, we will give here three (equivalent) definitions of this structure, beginning from the one which seems to be intuitively most clear and ending with the most operative one. The third definition gives a link between the remaining two.[3]

3.1a. Definition. Let $\mathcal{P}(X) = (\mathcal{P}, X, \pi, \mathcal{G})$ be a principal fibre bundle. A mapping $\sigma_p: \mathcal{T}_x(X) \to \mathcal{T}_p(\mathcal{P})$, for every $p \in \mathcal{P}$, $\pi(p) = x \in X$, is said to be a *connection* in $\mathcal{P}(x)$, if

[3]Here I closely follow Choquet-Bruhat et al. 1982, pp. 357–361.

(1) σ_p is linear,
(2) $\pi' \circ \sigma_p = \mathrm{id}$ on $\mathcal{T}_x(X)$,
(3) σ_p depends differentiably on p,
(4) $\sigma_{I\!R_g p} = I\!R' \sigma_p$, for every $g \in \mathcal{G}$.

The above conditions are chosen in such a way that the parallel transport made with the help of this connection is consistent with the structure of the fibre bundle as a differentiable and principal fibre. If $\mathbf{R}^1 \supset \mathcal{I} \to X$ is a curve $t \mapsto C(t)$, $t \in \mathcal{I}$, in X passing through a point $x_0 = \pi(p_0)$, $p_0 \in \mathcal{P}$, the parallel transport along the curve C is given by the curve $\tilde{C}: \mathbf{R}_1 \supset \mathcal{I} \mapsto \mathcal{P}$ which is defined through the following relationships[4],

$$\frac{d\tilde{C}(t)}{dt} = \sigma_p \frac{dC(t)}{dt},$$
$$\tilde{C}(0) = p_0, \quad \tilde{C}(t) = p.$$

The curve $\tilde{C}(t)$ in $\mathcal{P}$ is called the *horizontal lift* of the curve $C(t)$ in X. A horizontal lift $\tilde{C}(t)$ of a curve $C(t)$ determines a correspondence between every two fibres along $\tilde{C}(t)$, that is to say, it connects the bundle into an entirety. Condition (4) guarantees that two different horizontal lifts $\tilde{C}_1$ and $\tilde{C}_2$ of the same curve C in X, to two different points p_1 and p_2 of the same fibre $\pi^{-1}(x)$ over the point $x \in X$, are connected with each other with the help of the right action $I\!R_g$ of the structural group on the fibre $\pi^{-1}(x)$ in such a way that $I\!R_g$ transforms p_1 into p_2.

Let us define a *space of horizontal vectors* (or a *horizontal space*) at a point $p \in \mathcal{P}$ as

$$\mathcal{H}_p \equiv \{\sigma_p(\mathcal{T}_x(X))\}, \quad x = \pi(p).$$

A *space of vertical vectors* (or a *vertical space*) can be defined, independently of any connection, as the set of vectors which are tangent to a fibre, i.e.,

$$\mathcal{V}_p \equiv \{v \in \mathcal{T}_p(\mathcal{P}):\ \pi' v = 0\}, \quad p \in \mathcal{P}.$$

It is easily seen that the tangent space $\mathcal{T}_p(\mathcal{P})$ to the bundle at a point p can be decomposed into a direct sum of the horizontal and vertical spaces:

$$\mathcal{T}_p(\mathcal{P}) = \mathcal{H}_p \oplus \mathcal{V}_p,$$

i.e., for every $v \in \mathcal{T}_p(\mathcal{P})$, $v = v_H + v_V$, with $v_H \in \mathcal{H}_p$, $v_V \in \mathcal{V}_p$. Evidently, one has

$$\pi' \mathcal{H}_p = \mathcal{T}_x(X) \quad \text{and} \quad \pi' \mathcal{V}_p = 0, \quad x = \pi(p).$$

The first of these relations defines the isomorphism of the horizontal space $\mathcal{H}_p$ and the tangent space $\mathcal{T}_x(X)$, and suggests that $\mathcal{H}_p$ can be used to an alternative definition of the connection.

[4]In the following, magnitudes defined in a bundle will be denoted by the same symbols as the corresponding magnitudes in the base manifold but equipped with an additional tilde above the corresponding symbol.

3.1b. Definition. A *connection* in a principal fibre bundle $\mathcal{P}(X) = (\mathcal{P}, X, \pi, \mathcal{G})$ is a field of vector spaces $\mathcal{H}_p \subset \mathcal{T}_p(\mathcal{P})$ satisfying the following conditions:

(1) the mapping $\pi' : \mathcal{H}_p \to \mathcal{T}_x(X)$, $x = \pi(p)$, is an isomorphism,

(2) the field $\mathcal{H}_p$ depends differentiably on p,

(3) $\mathcal{H}_{I\!R_g p} = I\!R'_g \mathcal{H}_p$.

The transition to the third definition is the following. A correspondence between the spaces $\mathcal{H}_p$ and $\mathcal{V}_p$ suggests that the space $\mathcal{V}_p$ can be also used to define connection. This suggestion is correct. It turns out that vertical spaces $\mathcal{V}_p$ are isomorphic with the tangent space $\mathcal{T}_e(\mathcal{G})$ to the structural Lie group (as a manifold) at its neutral element e, i.e., with the Lie algebra G of the structural group $\mathcal{G}$. It is this isomorphism that leads to the third definition.

As is well-known, the space of all left invariant vector fields on a Lie group $\mathcal{G}$ (equipped with the Lie bracket operation) is the Lie algebra G of the Lie group $\mathcal{G}$. The Lie algebra G (as a vector space) of a Lie group $\mathcal{G}$ can be identified with a tangent space to the Lie group $\mathcal{G}$ at its neutral element e through the mapping $G \ni \vartheta \mapsto \vartheta(e) \in \mathcal{T}_e(\mathcal{G})$. On the other hand, to any element ϑ of the Lie algebra G of the structural group $\mathcal{G}$ there corresponds a vector field $\lambda(\vartheta)$ on $\mathcal{P}$ in the following way. If $\vartheta \in G$, then $\exp(\vartheta t)$ is a one-parameter subgroup of $\mathcal{G}$ which acts on $\mathcal{P}$ to the right. This allows us to draw the curve $p_t = I\!R_{\exp(\vartheta t)}(p)$ through every point $p \in \mathcal{P}$. Now, for any real-valued function f on $\mathcal{P}$ we define

$$\lambda(\vartheta)_p(f) = \frac{d}{dt}\left(f(p_t)\right)\Big|_{t=0}.$$

As can be easily seen, the vector $\lambda(\vartheta)_p$ is tangent to the fibre $\pi^{-1}(\pi(p))$ at p; it is, therefore, a vertical vector, $\lambda(\vartheta)_p \in \mathcal{V}_p$. The mapping $\lambda : G \to \mathcal{V}_p$ establishes a *canonical isomorphism* between the Lie algebra G of the structural group $\mathcal{G}$ and a vertical space $\mathcal{V}_p$, for every $p \in \mathcal{P}$. The vector field $\lambda(\vartheta)$ on $\mathcal{P}$ is called a *fundamental vector field*.

Now, a Lie algebra valued 1-form on $\mathcal{P}$,

$$\omega_p : \mathcal{T}_p(\mathcal{P}) \longrightarrow G,$$

can be defined in the following way: if $v \in \mathcal{T}_p(\mathcal{P})$ then $\omega_p(v) = \vartheta \in G$ such that $\lambda(\vartheta) = v_V \in \mathcal{V}_p$. Or, in other words, one demands that $\omega_p(v)$ should be equal to the vertical component of the vector v which, in turn, by the canonical isomorphism is identified with an element of the Lie algebra. The 1-form ω, called a *connection form*, determines a connection in the principal fibre bundle according to the following.

3.1c. Definition. A *connection* in a principal fibre bundle $\mathcal{P}(X) = (\mathcal{P}, X, \pi, \mathcal{G})$ is a 1-form ω on $\mathcal{P}$ satisfying the following conditions:

(1) For every $v \in \mathcal{V}_p$, $\omega_p(v) = \vartheta$ such that between v and ϑ there is a canonical isomorphism,

(2) ω_p depends differentiably on p,

(3) $\omega_{I\!R_g p}(I\!R'_g v) = \mathrm{Ad}(g^{-1})\omega_p(v)$.

Let us notice that the set of all vectors $v \in \mathcal{T}_p(\mathcal{P})$ for which $\omega_p(p) = 0$ determines the horizontal space $\mathcal{H}_p$. This fact allows one immediately to check the equivalence of the second and third definitions of connections. Condition (3) of the above definition is analogous to conditions (4) and (5) of the first and the second definition, correspondingly, but it is less transparent and requires a few words of comment.

The restriction of the form ω to a single fibre $\mathcal{G}_x = \pi^{-1}(x)$ gives

$$\omega(v_V)\big|_{\mathcal{G}_x} = \vartheta_V, \quad v_V \in \mathcal{T}(\mathcal{G}_x) \equiv \mathcal{V}_p$$

If we perform the mapping $s: \mathcal{G} \to \mathcal{G}_x$ such that $s(h) = p$ if and only if $p = I\!R_h p_0$, where $h \in \mathcal{G}$ and p_0 is a fixed point in $\mathcal{G}_x$, the form $\omega|_{\mathcal{G}_x}$ can be regarded as the Maurer–Cartan form of the Lie group $\mathcal{G}$.[5] It can be proved that for a Maurer–Cartan form one has

$$I\!R_g^* \omega(v) = \mathrm{Ad}(g^{-1})\omega(v),$$

where $\mathrm{Ad}(g)$; $g \in G$ is the so–called *adjoint representation* of the group $\mathcal{G}$; it is a representation in the Lie algebra G of this group (see Choquet–Bruhat et al. 1982, pp. 168–169).

Keeping in mind that the action $I\!R_g$ preserves the decomposition of a vector $v \in \mathcal{T}_p(\mathcal{M})$ into its horizontal component v_H and its vertical component v_V (see condition (3) of the second definition of connection), it is easy to compute the pull back of ω by $I\!R_g$

$$I\!R_g^*(\omega)v = \omega(I\!R_g' v) = \omega(I\!R_g' v_H + I\!R_g' v_V) = \omega(I\!R_g' v_V).$$

By treating ω as a Maurer–Cartan form on $\mathcal{G}$, one immediately obtains condition (3) of the third definition of connection.

To conclude this section, let us define parallel transport in an associated fibre bundle.

3.2. Definition. Let $\mathcal{E}(X) = (\mathcal{E}, X, \pi, \mathcal{G}, \mathcal{F})$ be an associated fibre bundle with a principal fibre bundle $\mathcal{P}(X) = (\mathcal{P}, X, \pi, \mathcal{G})$ by the representation $\rho: \mathcal{G} \to I\!L(\mathcal{F}, \mathcal{F})$. The *parallel transport* of an element $v_1 \in \mathcal{E}, \pi_1(v_1) = x_1$, along a curve C from the point $x_1 \in X$ to a point $x_2 \in X$, gives — from the definition — the element $v_2 \in \mathcal{E}$, $\pi_1(v_2) = x_2$ such that

$$\phi_i(v_2) = \rho(\Psi_i(p_2) \circ \Psi_i(p_i)^{-1})\phi_i(v_1),$$

where $p_2 \in \mathcal{P}$ is obtained by the parallel transport of the element $p_1 \in \mathcal{P}$, $\pi(p_1) = x_1$, in the principal fibre bundle $\mathcal{P}(X)$, along the curve C from the point x_1 to the point x_2. All other symbols are the same as in sections 1 and 2 with the exception that in symbols $\phi_{i,x}$ and $\Psi_{i,x}$, the index x was omitted for simplicity.

[5] Let us remember that the *Maurer–Cartan form of the Lie group* $\mathcal{G}$ is a form ω, defined on $\mathcal{G}$, with values in the Lie algebra G of the Lie group $\mathcal{G}$, such that $\omega(v_g) = \vartheta$, where $\vartheta = L_g^{-1} v_g \in \mathcal{T}_e\mathcal{G}$.

The parallel transport in an associated fibre bundle, defined in the above way, depends on neither the choice of p_1 in a fibre $\pi^{-1}(x_1)$ nor the choice of a trivialization of the fibre bundle.

4. Geometry of Principal Fibre Bundles

As we have seen, connection in a principal fibre bundle $\mathcal{P}(X)$ is a 1–form on $\mathcal{P}$ with values in a vector space, namely in the Lie algebra G of the structural group $\mathcal{G}$. Let us recall that a p–form on a differential manifold X with values in a vector space $\mathcal{V}$ is defined as

$$\varphi = \varphi^\alpha \otimes e_\alpha,$$

where φ^α is an ordinary p–form on X (with scalar values), and e_α are basis vectors in $\mathcal{V}$. It is assumed that $\mathcal{V}$ is a real and finite–dimensional space.

An *external derivative* of a p–form φ is defined to be

$$d\varphi = d\varphi^\alpha \otimes e_\alpha.$$

It can be easily checked that this definition does not depend on the choice of a basis in $\mathcal{V}$. Moreover, if $\mathcal{V}$ is a Lie algebra with the bracket operation $[v_\alpha, v_\beta]$ between every pair of vectors $v_\alpha, v_\beta \in \mathcal{V}$, then the bracket operation can be extended to forms assuming their values in $\mathcal{V}$. Let $\varphi = \varphi^\alpha \otimes e_\alpha$ and $\psi = \psi^\beta \otimes e_\beta$ be p–forms defined on X with values in $\mathcal{V}$. From the definition, we have

$$[\varphi, \psi] = (\varphi^\alpha \wedge \psi^\beta) \otimes [e_\alpha, e_\beta],$$

where $\wedge$ is the external product of forms. This definition is independent of the choice of a basis in $\mathcal{V}$.

Let us consider a principal fibre bundle $\mathcal{P}(X) = (\mathcal{P}, X, \pi, \mathcal{G})$ with a connection in it defined as a 1–form ω. Let

$$h\colon\ \mathcal{T}_p(\mathcal{P}) \longrightarrow \mathcal{H}_p$$

be a mapping defined by $v \mapsto v_H, v \in \mathcal{T}_p(\mathcal{P})$. Let also $\varphi = \varphi^\alpha \otimes e_\alpha$ be an r–form with values in a vector space $\mathcal{V}$ having a basis e_α. An *external covariant derivative* of an r–form φ is defined to be

$$D\varphi(v_1, \ldots, v_r, v_{r+1}) = d\varphi(hv_1, \ldots, hv_r, hv_{r+1}),$$

where, as usual, $d\varphi = d\varphi^\alpha \otimes e_\alpha$.

An r–form α on $\mathcal{P}(X)$ with values in a vector space $\mathcal{V}$ is called *horizontal* when $\alpha(v_1, \ldots, v_r) = 0$, if at least one of the vectors $v_1, \ldots, v_r$ is vertical.

4.1. Definition. An r–form α on $\mathcal{P}(X)$ with values in a vector space $\mathcal{V}$ is said to be a *form of the type* $(\rho, \mathcal{V})$, where ρ is the representation of the structural group $\mathcal{G}$ in the space $\mathcal{V}$, $\rho: \mathcal{G} \to \mathbb{L}(\mathcal{V}, \mathcal{V})$, if

$$\mathbb{R}_g^* \alpha = \rho(g^{-1})\alpha.$$

If this relation takes place the form α is said to be *equivariant* under the action of $\mathbb{R}_g$ with respect to the representation ρ. If additionally the r–form α is horizontal, it is called a *tensorial form of the type* $(\rho, \mathcal{V})$. Let us note that the type $(\rho, \mathcal{V})$ of a tensorial form informs us about the character of the tensor field in which, in the general case, the given form assumes its values; namely, the dimension of the space $\mathcal{V}$ is equal to the number of factors in the tensor product of tangent and cotangent spaces defining the given tensor, and the representation $\rho(g^{-1})$ conveys information about the properties of this tensor.

4.2. Definition. An external covariant derivative of a connection one–form ω on $\mathcal{P}(X)$ is said to be a *curvature form of the connection* ω and is denoted by Ω, i.e., $\Omega = D\omega$. Of course, Ω is a 2–form on $\mathcal{P}(X)$ with values in the Lie algebra G.

By using the relationship $\mathbb{R}_g^* \omega = \mathrm{Ad}(g^{-1}) \circ \omega$ (see sec. 3), one shows that the curvature form is a tensorial form of the type (Ad, G).

If $\Omega = D\Omega \equiv 0$ the connection is said to be *flat*.

4.3. Theorem. A connection form can be expressed with the help of the following equation, called the *Cartan* or *structural equation* (or *Cartan structural equation*)[6]

$$D(u, v) = d\omega(u, v) + [\omega(u), \omega(v)]. \quad \blacksquare$$

Before proving this theorem let us recall a useful formula expressing the external derivative of a form Θ,

$$d\Theta(u, v) = L_u \Theta(v) - L_v \Theta(u) - \Theta([u, v]),$$

where $L_u(v)$ is a Lie derivative of the vector v in the direction of the vector u. This formula can be proved by working it out in components.

Proof.

$$\begin{aligned} \Omega(u, v) = D\omega(u, v) &= d\omega(hu, hv) = d\omega(u - u_V, v - v_V) \\ &= d\omega(u, v) - d\omega(u, v_V) - d\omega(u_V, v) + d\omega(u_V, v_V) \\ &= d\omega(u, v) - d\omega(u_H, v_V) - d\omega(u_V, v_H) - d\omega(u_V, v_V). \end{aligned}$$

The second term of the middle row, $d\omega(u, v_V)$, has been replaced by the term $-d\omega(u_H + u_V, v_V)$, and analogously, the third term in this row. Two middle terms in the last row are equal to zero. Indeed, let us consider, for example, the second of them,

$$d\omega(u_H, v_H) = L_{u_V} \omega(v_H) - L_{v_H} \omega(v_V) - \omega[u_V, v_H].$$

[6] Sometimes the last term on the right hand side of the Cartan equation is multiplied by $\frac{1}{2}$; it depends on conventions assumed in the definition of external derivative.

Let us notice that $\omega(v_V) = \vartheta_V \in G$, and $\vartheta_V = \text{const}$. Therefore, the Lie derivatives vanish. However, the last term is a horizontal vector since

$$[u_V, v_H] = L_{u_V} v_H = \lim_{t \to 0} \frac{1}{t} (I\!R'_{g(t)} v_H - v_H),$$

that is to say, $d\omega(u_V, v_H) = 0$.

It remains to compute the term $d\omega(u_V, v_V)$. By using the same formula as before, one obtains

$$d\omega(u_V, v_V) = -\omega[u_V, v_V] = -[\omega(u_V), \omega(v_V)] = -[\omega(u), \omega(v)]. \quad \square$$

By choosing a basis (e_α) in the Lie algebra G, the Cartan equations can be expressed in coordinates

$$\Omega^\alpha = d\omega^\alpha + \frac{1}{2} c^\alpha{}_{\beta\gamma} \omega^\beta \wedge \omega^\gamma.$$

Of course, the coefficients $c^\alpha{}_{\beta\gamma}$ are structure constants of the Lie group $\mathcal{G}$. By differentiating, one obtains

$$d\Omega^\alpha = \frac{1}{2} c^\alpha{}_{\beta\gamma} d\omega^\beta \wedge \omega^\gamma - \frac{1}{2} c^\alpha{}_{\beta\gamma} \omega^\beta \wedge d\omega^\gamma.$$

From the last equation one can see that

$$D\Omega^\alpha(u, v, w) = d\Omega^\alpha(hu, hv, hw) = 0,$$

for any u, v, w. This formula is called *Bianchi identities*; it can be written in the following compact form: $D\Omega = 0$.

5. *Geometry of the Frame Bundle*

In applications we are often interested in the geometry of a base manifold and individual fibres over its points are treated as an "internal structure" of these points. The bundle manifold is, in such an approach, a kind of a "phase space" of internal states of base points. Such a "phase space of internal states" turns out to be a useful concept in investigating the geometric structure of the base manifold. This is the goal of the present section.

Let $\mathcal{P}(X) = (\mathcal{P}, X, \pi, \mathcal{G})$ be a principal bundle, $\tilde{\omega}$ a connection form in it, and f a smooth (C^r) local cross-section of the fibre bundle, i.e., a cross-section of a region $\pi^{-1}(U)$ of the bundle,

$$f\colon\ X \supset U \longrightarrow f(U) \subset \mathcal{P},$$
$$\pi \circ f = \mathrm{id}.$$

5.1. Definition. A *connection 1–form* on U with values in G is defined as the pull back of the connection form in $\pi^{-1}(U) \subset \mathcal{P}$ by the mapping f, i.e.,

$$\omega_x = (f^*\tilde{\omega})_x(u) = \omega_{f(x)}(f'u),$$

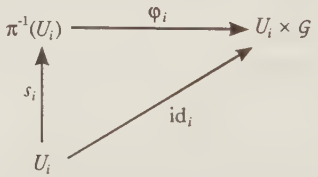
where $x \in U$, $u \in \mathcal{T}_x(X)$.

It can be proved that if a 1–form ω on $U \subset X$ is given, with values in G, and f is a smooth cross–section of the bundle $\pi^{-1}(U)$, then there exists the unique connection $\tilde{\omega}$ in $\pi^{-1}(U)$ such that $\omega = f^*\tilde{\omega}$. The proof consists in, constructing a 1–form $\tilde{\omega}$ at $p \in f(U)$ with the required properties, and extending it to the entire region $\pi^{-1}(U)$. Such a form at a point $p \in f(U)$ is given by

$$\tilde{\omega}_p(v) = \omega_x(\pi' v) + \vartheta_2,$$

where $v \in \mathcal{T}_p(\mathcal{P})$, $p = f(x)$, is a vector such that $v = v_1 + v_2$ and, from the definition, $v_1 = (f' \circ \pi')v$, i.e., v_1 is a horizontal vector and, consequently, v_2 is a vertical vector ($\pi' v_2 = 0$). The vector $\vartheta_2 \in G$ is obtained from the vector v_2 with the help of the canonical isomorphism between the Lie algebra of the structural group and the vertical spaces of the fibre bundle (see sec. 3). Extension of the form $\tilde{\omega}_p$ to all points of $\pi^{-1}(U)$ is made by using the action of the structural group $\mathcal{G}$ on $\pi^{-1}(U)$. (See Choquet–Bruhat et al. 1982, pp. 362–363).

The definition of a local connection on the base manifold X (i.e., of a connection on $U \subset X$) essentially depends on a local cross–section f of the bundle (i.e., on a cross–section of $\pi^{-1}(U)$). Among many such cross–sections one is privileged; let us denote it by s_i and call it a *cross–section canonically associated with the trivialization* (U_i, φ_i) of the fibre bundle. Let us define the mapping $\mathrm{id}_i : U_i \times \mathcal{G}$ by $x \mapsto (x, e)$, where e is the neutral element of the group $\mathcal{G}$; then, from the definition, $s_i = \varphi_i^{-1} \circ \mathrm{id}_i$. The following diagram will help to grasp the meaning of this definition:



5.2. Definition. The connection form $\omega_i = s_i^*\tilde{\omega}$ will be said to be a *connection form* (or simply a *connection*) *on a local trivialization* (U_i, φ_i). It plays an important role in physical applications.

With the help of a little bit more tedious proof (see Chcquet–Bruhat et al. 1982, pp. 364–365) one can show that if (U_i, φ_i) and (U_j, φ_j) are two local trivializations of the fibre bundle $\mathcal{P}(X)$, then two connections ω_i and ω_j on these trivializations, corresponding to the same connection $\tilde{\omega}$ in $\mathcal{P}$, i.e., $\omega_i = s_i^*\tilde{\omega}$ and $\omega_j = s_j^*\tilde{\omega}$ at a point $x \in U_i \cap U_j$, are related by the following formula,

$$\omega_{i,x} = \mathrm{Ad}(\gamma_{ij}^{-1}(x))\omega_{j,x} + (\gamma_{ij}^*\Theta_{\mathrm{MC}})_x,$$

where γ_{ij} are transition functions (see sec. 1), and Θ_{MC} is a Maurer–Cartan 1–form of the structural group $\mathcal{G}$ (see sec. 3, footnote 5).

A local curvature on a base manifold X is defined with the help of a similar procedure as a local connection on X.

5.3. Definition. Let (U_i, φ_i) be a local trivialization of a fibre bundle $\mathcal{P}(X)$ and $\tilde{\Omega}$ a curvature form on $\pi_i^{-1}(U) \subset \mathcal{P}$. A *2–form of curvature Ω_i on a local trivialization* (U_i, φ_i) is defined by

$$\Omega_i = s_i^* \tilde{\Omega}.$$

Two curvature forms Ω_i and Ω_j on two local trivializations (U_i, φ_i) and (U_j, φ_j) respectively corresponding to the same curvature form $\tilde{\Omega}$ on $\mathcal{P}$ at $x \in U_i \cap U_j$ are interrelated with the help of the following transformation,

$$\Omega_i(u, v) = \mathrm{Ad}(\gamma_{ij}(x)^{-1}) \Omega_j(u, v),$$

As it is directly seen, the Cartan equations in a local trivialization (U_i, φ_i) have the following form,

$$\Omega_i = d\omega_i + [\omega_i, \omega_i].$$

6. Applications to Gauge Theories

The number of applications of the fibre bundle geometry to physics quickly increases. The most spectacular, and perhaps the most important, of these applications were mentioned in the introduction to the present chapter. Applications to the theories of space–time, gravity, and cosmology will be studied in the following chapters. As a corollary to this chapter, let us enumerate (after Trautman 1980) some intersecting areas of the theory presented in the above sections and a geometrical formulation of gauge theories.

In the gauge theories one considers fibre bundles (both principal and associated) over space–time that plays the role of a base manifold. To construct a full gauge theory requires the selection of (1) a gauge group, (2) a particle type corresponding to the given gauge field, (3) a form of the field equations. The structure group of the fibre bundle is identified with the gauge group. In the Yang–Mills theories, connection forms on local trivializations (U_i, φ_i) of the principal bundle are what is called by physicists *gauge potentials*, curvature forms Ω_i on (U_i, φ_i) — *gauge fields*, and cross–sections of associated vector bundles — *matter fields*.

To be more precise, let us consider a principal fibre bundle $\mathcal{P}(\mathcal{M}) = (\mathcal{P}, \mathcal{M}, \pi, \mathcal{G})$ over space–time $\mathcal{M}$ and a vector bundle $\mathcal{E}(\mathcal{M}) = (\mathcal{E}, \mathcal{M}, \pi, \mathcal{G}, \mathcal{F})$ associated with $\mathcal{P}(\mathcal{M})$ by a representation $\rho: \mathcal{G} \to \mathbb{L}(\mathcal{F}, \mathcal{F})$. A *Higgs field of the type* $(\rho, \mathcal{F})$ is defined to be a cross–section $\psi: \mathcal{M} \supset \mathcal{N} \to \mathcal{E}$. When reading the literature one should pay attention to the fact that various authors accept different definitions; namely, some authors define a *particle of the type* $(\rho, \mathcal{F})$ as a mapping $\psi: \mathcal{P} \to \mathcal{F}$ satisfying

the condition $\bar{\psi}(vg) = \rho(g^{-1})\bar{\psi}(v)$ for every $p \in \mathcal{P}$ and $g \in \mathcal{G}$, and the *Higgs field* Ψ as the pull back of this mapping by the cross-section s_i of the principal fibre bundle, i.e., $\Psi = s_i^* \bar{\psi}: \mathcal{M} \to \mathcal{F}$ (fig. 3.3) (see Trautman 1980, pp. 296–297). The relationship between these two definitions of Higgs fields in a local map (U_i, φ_i) is given by $\psi = \phi_{i,x}^{-1} \circ \Psi$ (definition of $\phi_{i,x}$ in chapt. 3, sec. 1). Particle and Higgs fields are said to be *matter fields*.

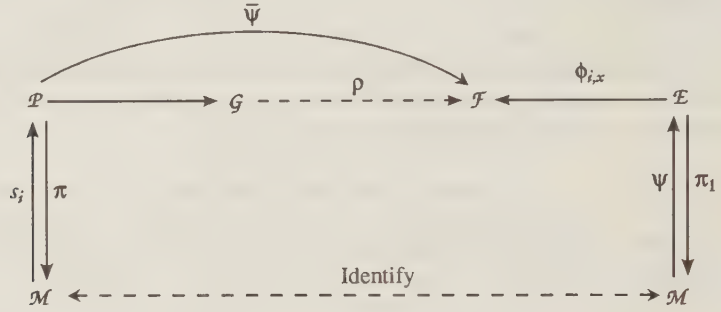


Fig. 3.3. Higgs field (one should remember that $\rho: \mathcal{G} \to \mathbb{L}(\mathcal{F}, \mathcal{F})$).

Another important notion is strictly connected with matter fields, namely the notion of the *spontaneous symmetry breaking*. Let us consider a particle of the type $(\rho, \mathcal{F})$

$$\bar{\psi}: \mathcal{P} \longrightarrow \mathcal{W} \subset \mathcal{F},$$

where the subspace $\mathcal{W}$ is an orbit of the Group $\mathcal{G}$ in $\mathcal{F}$, i.e., for any two points w_0, $w \in \mathcal{W}$ there is an element $g \in \mathcal{G}$ such that $w = \rho(g)w_0$. Let further $\mathcal{H}$ be an *isotropy group* (called also a *stability group*) of the element w_0, i.e.,

$$\mathcal{H} = \{g \in \mathcal{G}: \ \rho(g)w_0 = w_0\}.$$

Let us define

$$\mathcal{Q} \equiv \{p \in \mathcal{P}: \ \bar{\psi}(p) = w_0\}.$$

In such a case the fibre bundle $\mathcal{Q}(\mathcal{M})$ with $\mathcal{H}$ as its structural group is a reduction of the fibre bundle $\mathcal{P}(\mathcal{M})$ having the structural group $\mathcal{G}$ (see sec. 1). This reduction can be regarded as a geometric counterpart of the breaking of symmetries described by the group $\mathcal{G}$ to symmetries described by its subgroup $\mathcal{H}$.

It can be shown that the connection form ω_G (corresponding to the structural group $\mathcal{G}$), after the reduction of the fibre bundle $\mathcal{P}$ to the fibre bundle $\mathcal{Q}$, defines the connection form ω_H (corresponding to the subgroup $\mathcal{H}$) if and only if ω_H assumes its values in the Lie algebra H of the group $\mathcal{H}$ which in turn is equivalent to $D\bar{\psi} = 0$ (see Trautman 1980, pp. 297–298; 1981, pp. 409–413).

Let g be a metric of space–time $\mathcal{M}$ (the Minkowski metric is used most often), and $k_{ij} = k(e_i, e_j)$ be a non–singular metric in the Lie algebra G of the group $\mathcal{G}$ (k_{ij} is

invariant in the representation Ad), where (e_i) is a basis in G. The form $k_{ij} * \Omega^i \wedge \Omega^j$, invariant[7] with respect to the action of the group $\mathcal{G}$, can serve as a the Lagrangian of a gauge theory. In such a case sourceless field equations assume the form: $D * \Omega = 0$.

Bibliographical Note. Classical expositions on the theory of fibre bundles are Steenrod (1951) and Husemoller (1975); one could also use Sulanke and Wintgen (1977). The basic information can be found in every modern textbook of differential geometry, e.g. Kobayashi and Nomizu (1963) or Bishop and Crittenden (1964).

The first gauge theory comes from Weyl (1918); in this form he presented the electromagnetic theory within the framework of its unification with gravitation theory. Only twenty years later, the forgotten Weyl's idea was used by Schwinger (1951, 1953, 1962) in the theory of quantum fields. Great successes of gauge theories begin from the work by Yang and Mills (1954) in which the former ideas were applied to the theory of strong interactions. In the original version of the Yang–Mills theory, bosons transferring strong interactions turned out to be massless particles. This was the reason of the initial lack of interest in this theory. However, the crisis was efficiently overcome by Higgs (1964a, b, 1966) who proposed a mechanism of mass generation, called to–day a *Higgs mechanism*. It consists in a spontaneous symmetry breaking in interactions between the vacuum and the Higgs field. After generalizing, by Utiyama (1956), the Yang–Mills theory to other gauge symmetries, a dramatic expansion of these theories took place, crowned with the unification of electromagnetic and weak interactions by Weinberg (1967) and Salam (1968).

It was Lubkin (1963) who first noticed a connection between gauge theories and geometry of fibre bundles. The gauge theories were afterwards systematically developed in the language of fibre bundles by Trautman (1978, 1979, 1980); see also Dreschsler and Mayer (1977), Daniel and Viallet (1980), Eguchi et al. (1980), Choquet–Bruhat et al. 1982, pp. 401–408, Nash and Sen (1983, chapter 7), Trautman (1984). A popular paper by Berstein and Phillips (1984) presenting applications of fibre bundle methods to physics is recommended.

[7]Let us recall that the so–called *star operator* defines the isomorphism

$$*\colon \Lambda^p(X^n) \longrightarrow \Lambda^{(n-p)}(X^n)$$

by

$$\beta \longmapsto *\beta,$$

where the form $*\beta$, dual with respect to the p–form β, is a form such that $\tau(\alpha|\beta) = \alpha \wedge *\beta$ for every form $\alpha \in \Lambda^p(X^n)$, where τ is a volume element $\tau = \Theta^1 \wedge \cdots \wedge \Theta^n$, and $(\cdot|\cdot)$ denotes an inner product on Λ^p_x (see Choquet–Bruhat et al. 1982, pp. 294–295).

Chapter 4

The Frame Bundle and General Relativity

0. Introduction

One of the main reasons for going from special to general relativity was for Einstein the idea to extend the equivalence of all inertial reference frames to all frames of reference. As is well–known, such an equivalence, in general, can be achieved only locally. If a basis at a given point of space–time is to be considered as the mathematical counterpart of a local reference frame, then it is no wonder that the bundle of bases (*frame bundle*) over a space–time manifold provides a suitable mathematical environment to formulate the general theory of relativity.

The frame bundle is a principal fibre bundle but, being a "concrete" bundle, it has a richer structure than "abstract" bundles, which were studied in the preceding chapter and which are used to geometrize gauge theories (of the Yang–Mills type). The richness of the frame bundle structure and its specific properties are strictly connected with the existence of the soldering form θ in this bundle. This form — as its name picturesquely suggests — solders the frame bundle with its base manifold (space–time) which reflects the fact that the structure of the bundle is, in principle, determined by the structure of the base manifold. From the point of view of general relativity it is a natural property since, in this theory, gravitational field is "reduced" to the space–time geometry.

A fibre $\pi^{-1}(x)$ of the frame bundle over a point x of space–time forms a set of all bases at the point x, i.e., the set of all possible frames of references at x. One could say that in the case of the frame bundle over space–time, an "internal space" of an event x is equipped with a structure which enables one to describe all phenomena occurring in the neighbourhood of x, *with respect* to any frame of reference with its origin at x. In this sense, the idea of relativity is incorporated into the rich concept of the frame bundle over space–time.

It turns out, however, that the structure of the internal space of events of the type "any possible frame at x" is too general (or too vague). As we know from chapter 2, physics can develop on a manifold only if this manifold carries a Lorentz metric. On the other hand, the appearance of a Lorentz metric on the base manifold reduces the frame bundle to the subbundle of orthonormal frames, or, in other words, it corresponds to a spontaneous breaking of symmetries (described by the $\mathcal{GL}(\mathbf{R}^4)$ group) in the frame bundle into symmetries of the Lorentz group. The very possibility of physics enforces such a spontaneous symmetry breaking. This is reflected in

the structure of internal spaces in such a way that now this structure admits only pseudoorthogonal local frames. Experimental results tell us that physics distinguishes exactly this realization of the principle of relativity.

In the present chapter the following topics are discussed: geometry of the frame bundle (sec. 2 and 3), the reduction of the frame bundle to the subbundle of orthonormal frames (sec. 4). All these are only a preparation for sec. 5, in which the mathematical structure of general relativity is discussed; we will also focus on the problem of how the frame bundle structure leads to the correct choice of connection.

In *Comments and Remarks* (sec. 6) some of possible generalizations of general relativity (the so–called connection–metric theories) are mentioned. This time, the *Bibliographical Note* at the end of the chapter constitutes its important part; it urges the reader not to be satisfied with a rather compact presentation of general relativity given here but to reach for at least some of available textbooks and monographs.

1. Geometry of the Frame Bundle (continuation)

The subject matter of the present and following sections is the bundle of frames $\mathcal{F}(X^n) = (\mathcal{F}, X^n, \pi, \mathcal{GL}(\mathbf{R}^n))$ (see chapt. 3, example 2.2). Let us notice that a frame at a point $x \in X^n$ can be also defined[1] as a non–singular linear mapping $\rho_x: \mathbf{R}^n \to \mathcal{T}_x(X^n)$ defined by $\rho_x(u_1, \ldots, u_n) = u \in \mathcal{T}_x(X^n)$; the numbers $\rho_x(u_1, \ldots, u_n)$ are components of the vector u in the frame ρ_x. It can be proved that every fibre bundle of the form $\mathcal{P}(X^n) = (\mathcal{P}, X^n, \pi, \mathcal{GL}(\mathbf{R}^n))$ is isomorphic[2] with the frame of bundles if an only if one can define a *soldering form* in the bundle $\mathcal{P}(X^n)$. Let us define it and explore some of its properties.

1.1. Definition. Let the mapping

$$\Phi_p: \mathcal{T}_x(X^n) \longrightarrow \mathbf{R}^n,$$

where $p = (x, \rho_x) \in \mathcal{F}(X^n)$ is a frame (basis) ρ_x (with the basis vectors (e_i)) at $x \in X^n$, be given by

$$\Phi_p(u) = \theta^i(u), \quad u \in \mathcal{T}_x(X^n),$$

where θ^i is a dual basis with respect to (e_i). In other words, the mapping Φ_p coordinates to every $u \in \mathcal{T}_x(X^n)$ its components in the frame ρ_x. If one prefers to define a frame as $\rho_x: \mathbf{R}^n \to \mathcal{T}_x(X^n)$, then $\Phi = \rho_x^{-1}$. A one–form

$$\theta: \mathcal{F}(X^n) \longrightarrow \mathbf{R}^n$$

defined by

$$\theta_p(v) = \Phi_p(\pi' v)$$

[1] Provided that a base in $\mathbf{R}^n$ has been defined.

[2] In the category of principal fibre bundles.

is called a *soldering form* (or a *canonical form*) on $\mathcal{F}(X^n)$. One speaks also of the *soldering* of the manifold X^n with the bundle of frames. (Since in the present section we are interested principally only in forms defined on frame bundle, we shall not distinguish them from those on the base manifold by the tilde.)

The soldering form can be expressed in coordinates

$$(\theta_p(v))^\gamma = (a^{-1})^\gamma{}_\beta dx^\beta(\pi' v).$$

Indeed, if $\partial_\alpha = \partial/\partial x^\alpha$, $(\rho_x)_\alpha = a^\beta{}_\alpha \partial_\beta$, $\theta = \theta^\alpha e_\alpha$, where (e_α) is the natural basis in $\mathbf{R}^n$, then

$$\begin{aligned}\theta_p(v) &= \Phi_p(\pi' v) = \Phi_p(dx^\beta(\pi' v)\partial_\beta)\\ &= \Phi_p(dx^\beta(\pi' v)(a^{-1})^\gamma{}_\beta a^\delta{}_\gamma \partial_\delta)\\ &= dx^\beta(\pi' v)(a^{-1})^\gamma{}_\beta e_\gamma.\end{aligned}$$

The soldering form θ on $\mathcal{F}(X^n)$ is a tensorial form of the type $(\mathrm{id}, \mathbf{R}^n)$ (see chapt. 3, definition 4.1). Indeed, from its definition θ is a horizontal form. Moreover, if $g \in \mathcal{GL}(\mathbf{R}^n)$, $\mathrm{id}(g) = g$, $v \in \mathcal{T}_p(\mathcal{F}(X^n))$, then

$$\begin{aligned}(I\!R_g^* \theta)_p(v) &= \theta_p(I\!R_g' v) = \Phi_p(\pi'(I\!R_g' v))\\ &= \Phi_{I\!R_g p}(\pi' v) = g^{-1}\Phi_p(\pi' v) = g^{-1}\theta_p(v).\end{aligned}$$

The one–before–last equality follows from the fact that if $p = (x, \rho_x) \mapsto I\!R_g p = (x, \rho_x')$ then $\rho_x \mapsto \rho_x' = g^{-1}\rho_x$.

1.2. Definition. A connection in the bundle of frames is called a *linear connection*. The 2–form $\Theta = D\theta$ is said to be a *torsion form* of the linear connection in $\mathcal{F}(X^n)$.

Now, we shall prove the following.

1.3. Theorem. The torsion form can be expressed with the help of the following equation, called the *Cartan equation* or the *structure equation* (or the *Cartan structure equation*) *for the torsion*,

$$\Theta(u, v) = d\theta(u, v) + [\omega(u), \theta(v)],$$

where $[\omega(u), \theta(v)] \equiv \omega(u)\theta(v) - \omega(v)\theta(u)$, and $\omega(u)\theta(v)$ denotes the action of $\omega(u)$, which is an element of the Lie algebra $GL(\mathbf{R}^n)$ of the group $\mathcal{GL}(\mathbf{R}^n)$, on $\theta(v) \in \mathbf{R}^n$. ∎

Proof.

$$\begin{aligned}\Theta(u, v) \equiv D\theta(u, v) &= d\theta(u_H, v_H)\\ &= d\theta(u, v) - d\theta(u, v_V) - d\theta(u_V, v) + d\theta(u_V, v_V),\\ d\theta(u_V, v) &= L_{u_V}\theta(v) - \theta([u_V, v]),\end{aligned}$$

since $L_v\theta(u_V) = 0$ as θ is a horizontal form, but

$$\theta([u_V, v]) = i_{[u_V, v]}\theta = [L_{u_V}, i_v]\theta.$$

We have used the notation $i_v\alpha = \alpha(v)$, where α is a 1–form and v a contravariant vector (i_v can be thought of as an internal product operation), and the following equality,

$$[L_v, i_w]\alpha = i_{[v,w]}\alpha$$

which can be easily checked (by direct computation in coordinates). By joining together the two expressions above after straightforward manipulation, one obtains

$$d\theta(u_V, v) = L_{u_V}\theta(v) - L_{u_V} i_v\theta + i_v L_{u_V}\theta = i_v L_{u_V}\theta.$$

Let us note that

$$\left.\frac{dg(t)}{dt}\right|_{t=0} = \omega(u)$$

(this follows from the definition of connection as a form with its values in the Lie algebra of the structural group). By using the definition of Lie derivative and remembering that $I\!R_g$ acts "vertically", we arrive at

$$i_v L_{u_V}\theta = i_v \frac{d}{dt}(I\!R^*_{g(t)}\theta)\Big|_{t=0} = i_v \frac{d}{dt}(g^{-1}\theta)\Big|_{t=0} = -\omega(u)\theta(v),$$

that is to say,

$$d\theta(u_V, v) = -\omega(u)\theta(v).$$

A similar procedure leads to

$$d\theta(u, v_V) = \omega(u)\theta(v).$$

The term $d\theta(u_V, v_V)$ vanishes since θ is a horizontal form.

By collecting together all the above results we obtain the Cartan structure equation. □

There exists a simple relationship between torsion and curvature forms. By taking external covariant derivative of the torsion form, we find

$$D\Theta = \Omega \wedge \theta.$$

The easiest way to do this is to express the second term of the right–hand side of the Cartan equation as an external product. Let us recall that the external product of a p–form α and a q–form β is defined to be

$$(\alpha \wedge \beta)(v_1, \ldots, v_{p+q}) = \frac{1}{p!\,q!}\sum_{\pi}(\operatorname{sign}\pi) \times \pi(\alpha(v_1, \ldots, v_p)\beta(v_{p+1}, \ldots, v_{p+q})),$$

where π denotes permutation of indices. One should note that sometimes, in place of $1/(p!\,q!)$, the factor $1/(p+q)!$ is used in the above formula; in such a case antisymmetrization makes the factor $1/2$ to appear.

2. Geometry of the Base Manifold

First we will present a method to construct forms in a base manifold X^n (with values in the space of tensor fields of a given type), canonically associated with tensorial forms (of the type $(\rho, \mathbf{R}^{n^q})$) in the bundle of frames $\mathcal{F}(X^n)$. Afterwards, by using this method we shall construct forms on X^n, canonically associated with the soldering form $\tilde{\theta}$, the torsion form $\tilde{\Theta}$, and the connection form $\tilde{\omega}$ in $\mathcal{F}(X)$. In this way, the geometry of the base manifold will be determined in terms of the frame bundle over X^n.

Now, we will generalize to the tensorial case the mapping Φ_p, defined at the beginning of the preceding section. Let us consider a manifold X^n and let Φ_p, $p = (x, \rho_x)$, $x \in X^n$, map any tensor on X^n into its coordinates with respect to the frame ρ_x, i.e.,

$$\Phi_p\colon \otimes^s \mathcal{T}_x^*(X^n) \otimes^t \mathcal{T}_x(X^n) \longrightarrow \mathbf{R}^{n^q},$$

where $q = s + t$. Let us also consider $v_i \in \mathcal{T}_p(\mathcal{F}(X^n))$ and $u_i \in \mathcal{T}_x(X^n)$ such that $u_i = \pi'(p)v_i$.

2.1. Definition. An r–form α is said to be *canonically associated* with an r–form $\tilde{\alpha}$ defined on $\mathcal{F}(X^n)$ if

$$\alpha_x(u_1, \ldots, u_r) = \Phi_p^{-1}(\tilde{\alpha}(v_1, \ldots, v_r)),$$

where $x = \pi(p)$, and $\tilde{\alpha}$ is a tensorial form of the type $(\rho, \mathbf{R}^{n^q})$ with values in $\mathbf{R}^{n^q}$.

We shall show that this definition is independent of the choice of v_i and of the choice of the frame ρ_x.

Indeed, (1) it does not depend on the choice of v_i. Let v_i and v_i' be any vectors such that $u_i = \pi'(p)v_i = \pi'(p)v_i'$; then

$$\Phi_p^{-1}(\tilde{\alpha}(v_i)) - \Phi_p^{-1}(\tilde{\alpha}(v_i')) = \Phi_p^{-1}(\tilde{\alpha}(v_i - v_i')) = 0,$$

since $\tilde{\alpha}$ is a tensorial form (and consequently horizontal) and the difference $v_i - v_i'$ is a vertical vector.

(2) The definition does not depend on the choice of the frame ρ_x. Let $p = (x, \rho_x)$ and $p' = (x, \rho_x')$; in such a case there exists $g \in \mathcal{GL}(\mathbf{R}^n)$ such that $p' = I\!R_g p$. Let further $v_i' = I\!R_g' v_i$. Since $I\!R_g'$ acts vertically, we have $\pi' v_i' = \pi v_i = u_i$. Therefore, we obtain

$$\tilde{\alpha}_{p'}(v_i') = (I\!R_g^* \tilde{\alpha})_p(v_i) = \rho(g^{-1})\tilde{\alpha}_p(v_i)$$

and $\Phi_{p'}^{-1} = \Phi_p^{-1}\rho(g)$ on the strength of the definition of Φ_p. From the last two formulae it follows the independence of the definition of the choice of p. □

According to the above recipe the soldering 1–form $\tilde{\theta}$ on $\mathcal{F}(X^n)$ defines a 1–form θ on X^n. Since $\tilde{\theta}$ is a tensorial form of the type $(\mathrm{id}, \mathbf{R}^n)$, θ assumes its values in the space of contravariant vector fields. From the definitions of $\tilde{\theta}$ and θ we have

$$\theta_x(u) = \Phi_p^{-1}\tilde{\theta}(v) = \Phi_p^{-1}\Phi_p \pi' v = \pi' v = u,$$

and consequently θ defines the tensor field $\theta^{\alpha}_{\beta} = \delta^{\alpha}_{\beta}$.

Let us notice that the connection form in the principal bundle, being a non-tensorial form (it is not horizontal), does not define any form on the base manifold in the way discussed above. As a local connection form on the base manifold X^n we must employ the form $\omega_i = s_i^* \tilde{\omega}$ where s_i is the cross-section canonically associated with the trivialization (U_i, φ_i) (see chapt. 3, sec. 5). Having in mind future applications, let us compute

$$\omega_i(\pi' v) = s_i^* \tilde{\omega}(\pi' v) = \tilde{\omega}(s_i' \pi' v) = \omega(v),$$
$$v \in \mathcal{T}_p(\mathcal{F}(X^n)).$$

Remembering that the form ω_i assumes its values in the Lie algebra of the structural group of the bundle, we can also write $\omega(\pi' v) = \omega^{\alpha}{}_{\beta}(\pi' v)\mathcal{E}^{\beta}{}_{\alpha}$, where $(\mathcal{E}^{\beta}{}_{\alpha})$ is a basis in $\mathcal{GL}(\mathbf{R}^n)$ and $\omega^{\alpha}{}_{\beta}(\pi' v)$ is formally treated as a numerical matrix. The connection form $\tilde{\Omega}$ on $\mathcal{F}(X^n)$ is a tensorial 2-form of the type $(\mathrm{Ad}, \mathbf{R}^{n^2})$. Therefore, it defines the 2-form Ω on X^n

$$\Omega_x(\pi' v_1, \pi' v_2) = \Phi_p^{-1}(\tilde{\Omega}(v_1, v_2)).$$

From the type of the form $\tilde{\Omega}$ one sees that Ω assumes its values in the space of 1-covariant and 1-contravariant tensor fields. By using the structure equation for curvature, we compute

$$\Omega(\pi' v_1, \pi' v_2) = d\omega(\pi' v_1, \pi' v_2) + [\omega(\pi' v_1), \omega(\pi' v_2)],$$

or in coordinates

$$\begin{aligned}\Omega &= d\omega^{\alpha}{}_{\beta}(\pi' v_1, \pi' v_2) + \omega^{\alpha}{}_{\beta}(\pi' v_1)\omega^{\gamma}{}_{\delta}(\pi' v_2)[\mathcal{E}^{\beta}{}_{\alpha}, \mathcal{E}^{\delta}{}_{\gamma}] \\ &= d\omega^{\alpha}{}_{\beta}(\pi' v_1, \pi' v_2) + (\omega^{\alpha}{}_{\gamma} \wedge \omega^{\gamma}{}_{\beta})(\pi' v_1, \pi' v_2)\mathcal{E}^{\beta}{}_{\alpha}.\end{aligned}$$

In the last step we have used the equality $[\mathcal{E}^{\beta}{}_{\alpha}\mathcal{E}^{\delta}{}_{\gamma}] = \delta^{\beta}_{\gamma}\mathcal{E}^{\delta}{}_{\alpha} - \delta^{\delta}_{\alpha}\mathcal{E}^{\beta}{}_{\gamma}$. And finally

$$\Omega^{\alpha}{}_{\beta} = d\omega^{\alpha}{}_{\beta} + \omega^{\alpha}{}_{\gamma} \wedge \omega^{\gamma}{}_{\beta}.$$

It is the very well-known form from the standard expositions of differential geometry, *(Cartan) structural equation for the curvature form on the (base) manifold.*

The torsion form $\tilde{\Theta}$ on $\mathcal{F}(X^n)$ is a tensorial form of the type $(\mathrm{id}, \mathbf{R}^n)$, therefore it defines the 2-form Θ on X^n

$$\Theta_x(\pi' v_1, \pi' v_2) = \Phi_p^{-1}(\tilde{\Theta}(v_1, v_2)),$$

which assumes its values in the space of contravariant tensor fields. From the structure equation for the torsion form we obtain

$$\begin{aligned}\Theta^{\alpha}(\pi' v_1, \pi' v_2) = {} & d\theta^{\alpha}(\pi' v_1, \pi' v_2) + \omega^{\alpha}{}_{\beta}(\pi' v_1)\theta^{\beta}(\pi' v_2) \\ & - \omega^{\alpha}{}_{\beta}(\pi' v_2)\theta^{\beta}(\pi' v_1),\end{aligned}$$

where (θ^β) should be treated as a basis in $\mathcal{T}_x^*(X^n)$. This equation can be put into a more condensed form

$$\Theta^\alpha = d\theta^\alpha + \omega^\alpha{}_\beta + \theta^\beta.$$

If is the well-known *structural equation for the torsion form on the (base) manifold.*

3. Geometry of the Base Manifold (continuation)

In order to make the reader's work easier, we now collect the more important formulae of the differential geometry on the base manifold (in coordinates), without proof. A detailed discussion (with proofs) can be found in standard text-books on modern differential geometry (e.g., Choquet–Bruhat et al. 1982, chapt. 5; O'Neill 1983, chapt. 3; Bishop and Goldberg 1980, chapt. 5; Kobayashi and Nomizu 1963, passim) or in the more extensive presentations of general relativity (e.g., Misner et al. 1973, part. III; Weinberg 1972, chapt. 6). All tensors and forms appearing in the present section are defined on the base manifold.

The structure equations for torsion and curvature, correspondingly, can be written in the form

$$\Theta^i = d\theta^i + \omega^i_l \wedge \theta^l = \frac{1}{2} T^i{}_{kl} \theta^k \wedge \theta^l, \tag{3.1}$$

$$\Omega^j_i = d\omega^j_i + \omega^j_m \wedge \omega^m_i = \frac{1}{2} R^j{}_{ikl} \theta^k \wedge \theta^l, \tag{3.2}$$

where $T^i{}_{kl}$ and $R^j{}_{ikl}$ are *torsion and curvature tensors* respectively. Let (e_i) be a *moving frame* and (θ^j) a basis dual to (e_i). One has $[e_k, e_l] = c^i{}_{kl} e_i$; $c^i{}_{kl}$ are called *structure coefficients of the moving frame* (e_i). *Connection coefficients* $\gamma^j{}_{kl}$ are defined by the formula $\nabla e_l = \gamma^j{}_{kl} \theta^k \otimes e_j$, where ∇ denotes covariant derivative. The torsion and curvature tensors in local coordinates assume the following forms,

$$T^i{}_{kl} = T(\theta^i, e_k, e_l) = \gamma^i{}_{kl} - \gamma^i{}_{lk} - c^i{}_{kl}, \tag{3.3}$$

$$R_i{}^j{}_{kl} = R(e_i, \theta^j, e_k, e_l) = e_k(\gamma^j{}_{li}) - e_l(\gamma^j{}_{ki}) + \gamma^j{}_{km}\gamma^m{}_{li} - \gamma^j{}_{lm}\gamma^m{}_{ki} - c^m{}_{kl}\gamma^j{}_{mi}. \tag{3.4}$$

In the natural basis $(\partial/\partial x^i)$, for which $[\partial/\partial x^i, \partial/\partial x^j] = 0$, formulae (3.3) and (3.4) are reduced to

$$T^i{}_{kl} = \Gamma^i{}_{kl} - \Gamma^i{}_{lk}, \tag{3.5}$$

$$R_i{}^j{}_{kl} = \partial_k \Gamma^j{}_{li} - \partial_l \Gamma^j{}_{ki} + \Gamma^j{}_{km}\Gamma^m{}_{li} - \Gamma^j{}_{lm}\Gamma^m{}_{ki}, \tag{3.6}$$

where $\Gamma^i{}_{kl}$ are connection coefficients in the natural basis, called *Christoffel symbols*. The following formulae are valid in any local coordinates,

$$\sum_{(jli)} R_i{}^k{}_{jl} = \sum_{(jli)} (\nabla_j T^k{}_{li} - T^m{}_{ji} T^k{}_{ml}), \tag{3.7}$$

$$\sum_{(jkl)} \nabla_j R_i{}^m{}_{kl} = \sum_{(jkl)} T^h{}_{kj} R_i{}^m{}_{hl}. \tag{3.8}$$

The summation should be done after cyclic permutation (jli) of the three indices. Formulae (3.8) are called *Bianchi identities.*

4. The Fibre Bundle of Orthonormal Frames

4.1. Definition. Let us consider the following mapping,

$$\mathcal{F}\colon \mathcal{P}_1(X_1) = (\mathcal{P}_1, X_1, \pi_1, \mathcal{G}_1) \longrightarrow \mathcal{P}_2(X_2) = (\mathcal{P}_2, X_2, \pi_2, \mathcal{G}_2);$$

$\mathcal{F}$ is called a *bundle homomorphism* if the bundles $\mathcal{P}_1(X_1)$ and $\mathcal{P}_2(X_2)$ are equivalent (see chapt. 3, definition 1.4) and if there is a homomorphism $\phi\colon \mathcal{G}_1 \to \mathcal{G}_2$ such that $\mathcal{F}(I\!R_{g_1} p_1) = I\!R_{\phi(g_1)} \mathcal{F}(p_1)$, $p_1 \in \mathcal{P}_1$, $g_1 \in \mathcal{G}$. A bundle homomorphism will be denoted by $(\mathcal{F}, \phi)$.

4.2. Theorem. If between $\mathcal{P}_1(X_1)$ and $\mathcal{P}_2(X_2)$ there is a bundle homomorphism, then a connection on $\mathcal{P}_1(X_1)$ uniquely defines the connection on $\mathcal{P}_2(X_2)$. ■

Proof. Let us assume that $\mathcal{H}_{p_1} \subset \mathcal{T}_{p_1}(\mathcal{P}_1)$ defines a connection in $\mathcal{P}_1(X_1)$ (see definition 3.1b, chapter 3). Then $\mathcal{H}_{p_2} \equiv I\!R_{g_2}(\mathcal{F}'(p_1)), \mathcal{H}_{p_2} \subset \mathcal{T}_{p_2}(\mathcal{P}_2)$ defines a connection in $\mathcal{P}_2(X_2)$. From the properties of $I\!R_g$ and of homomorphism ϕ it follows that the definition of $\mathcal{H}_p$ does not depend on the choice of p_1 and p_2. (See Choquet–Bruhat et al. 1982, p. 380). □

Let $\mathcal{P}(X) = (\mathcal{P}, X, \pi, \mathcal{G})$ be a reduction of the frame bundle $\mathcal{F}(X)$ (see chapt. 3, sec. 1). The inclusions $\mathcal{P}(X) \subset \mathcal{F}(X)$ and $\mathcal{G} \subset \mathcal{GL}(\mathbf{R}^n)$ define a bundle homomorphism. From this and from theorem 4.2 it follows that a connection in $\mathcal{P}(X)$ uniquely defines the (linear) connection in $\mathcal{F}(X)$.

Let X^n be a Riemann manifold. A basis (e_i) on X^n is said to be an *orthonormal frame* if $(e_i|e_j) = \delta_{ij}$. Having an orthonormal frame (e_j) on X^n, we can obtain from it any other orthonormal frame $(e_{j'})$ with the help of the transformation $\bar{e}_{i'} = a^j_{i'} e_j$, where $a^j_{j'}$ is an element of the group $\mathcal{O}(n)$ of orthogonal matrices $n \times n$. There is, therefore, a natural bijection between the set of orthonormal frames on X^n and the group $\mathcal{O}(n)$.

Let us denote the set of pairs (x, τ_x), where $x \in X^n$ and τ_x is an orthonormal frame at x, by $\mathcal{O}(X^n)$. The bundle of orthonormal frames $\mathcal{O}(X^n) = (\mathcal{O}(X^n), X^n, \pi_1, \mathcal{O}(n))$ can be defined in a natural way; evidently, it is a reduction of the principal fibre bundle of (linear) frames $\mathcal{P}(X^n) = (\mathcal{P}, X^n, \pi, \mathcal{GL}(\mathbf{R}^n))$. On the strength of theorem 4.2 a connection in $\mathcal{O}(X^n)$ uniquely determines a connection on $\mathcal{P}(X^n)$ which is called the *metric connection*. Since an orthonormal basis, when parallelly transported with the help of the metric connection, remains an orthonormal frame, the covariant derivative of the metric tensor on X^n vanishes. Direct computation from (3.1) shows that the torsion of the metric connection is also equal to

zero. By using elementary tools of the Riemann geometry one can prove that, in any Riemann space, there is only one connection with these properties; it is also called the *Riemann connection*.

Let (F, ϕ) be a reduction of the bundle $\mathcal{F}(X)$ with the structural group $\mathcal{G}$ to a bundle $\mathcal{P}(X^n)$ with a structural group $\mathcal{G}_1 \subset \mathcal{G}$. The opposite question arises: does a connection in $\mathcal{F}(X^n)$ uniquely determine a connection in $\mathcal{P}(X^n)$? In general, since the mapping F need not be a surjection, the answer is negative. However, if X^n is a Lorentz (pseudoriemannian) manifold and (F, ϕ) reduction of $\mathcal{F}(X^n)$ to the bundle of orthonormal frames $\mathcal{O}_k(X^n)$, where k is an index of the manifold X^n (see chapt. 2, sec. 1 and 3), then the above question chooses one and only one connection in $\mathcal{O}_k(X^n)$ such that the torsion of this connection vanishes. To be more precise we formulate the following.

4.3. Definition. Let $F: \mathcal{P}_1 \to \mathcal{P}$, $\phi: \mathcal{G}_1 \to \mathcal{G}$ be inclusions defining a reduction of a principal fibre bundle $\mathcal{P}(X)$ with a structural group $\mathcal{G}$ to a fibre bundle $\mathcal{P}_1(X)$ with a structural group $\mathcal{G}_1$. One says that a connection ω in $\mathcal{P}(X)$ *can be reduced* to $\mathcal{P}_1(X)$ if there exists a connection ω_1 in $\mathcal{P}_1(X)$ such that $\omega_1 = F^*\omega$.

If a connection ω in $\mathcal{P}(X)$ can be reduced to $\mathcal{P}_1(X)$, then values of the connection form ω_1 belong to a Lie subalgebra G_1 of the Lie algebra G of the structural group $\mathcal{G}$, and the connection in $\mathcal{P}(X)$ uniquely determines a connection in $\mathcal{P}_1(X)$ (see Sulanke and Wintgen 1977, p. 171). As we have mentioned, in general this is not true. However, we have

4.4. Theorem. There exists one and only one linear connection ω in the frame bundle $\mathcal{F}(X^n)$ which can be reduced to the bundle of pseudoorthonormal frames $\mathcal{O}_k(X^n)$, being the reduction of $\mathcal{F}(X^n)$ to the pseudoorthogonal group $\mathcal{O}(n, k)$, such that the torsion Θ of ω is equal to zero. ■

Proof consists in, first, showing that the linear connection in $\mathcal{F}(X^n)$ can be reduced to the bundle $\mathcal{O}_k(X^n)$ if and only if $Dg = 0$, where g is a Lorentz metric (with index k on X^n). Then, by using the relationship $Dg = 0$, one constructs the unique connection in $\mathcal{O}_k(X^n)$. If turns out that the coefficients corresponding to this connection can be transformed to the well-known form of the Christoffel symbols. The detailed proof would require several additional theorems with their corresponding proofs which, in turn, would mean reproducing a large part of a differential geometry textbook. The interested reader should be referred, for instance, to the book by Sulanke and Wintgen (1972, chapt. 2, sec. 11). □

Let us stress that theorem 4.4 distinguishes exactly one connection in $\mathcal{O}_k(X^n)$; it will be called a *Lorentz connection* (the same name will also be used for the corresponding connection in X^n). Let us notice that, on the strength of theorem 4.2, the Lorentz connection in $\mathcal{O}_k(X^n)$ uniquely determines the linear connection in $\mathcal{F}(X^n)$.

In the light of the above considerations it can be said that the existence of a Lorentz structure in X^n breaks down the full $\mathcal{GL}(\mathbf{R}^n)$ symmetry of the frame bundle to the symmetry $\mathcal{O}(n, k)$ which will be also called the *Lorentz symmetry*.[3] In this

[3]It can be proved that the Lorentz structure (with index k) on a manifold X^n is equivalent to

sense, one can think that the Lorentz structure (or Lorentz metric) behaves similarly as a Higgs field in gauge theories (see chapt. 3, sec. 6).[4]

5. General Relativity

The mathematical theory developed in chapters 3 and 4 essentially allows one to enrich the metric model of the space-time presented in chapt. 2 (especially sec. 2). We have seen how the important role of enriching the manifold structure is played by connection. From a logical point of view, one should determine directly which connection one has in mind when speaking of the pair $(\mathcal{M}, g)$, and consequently the pair $(\mathcal{M}, g)$, as a model of space–time, should be replaced by the triple $(\mathcal{M}, \omega, g)$, where ω is a connection in the manifold $\mathcal{M}$. Every choice of connection leads to a different theory of space–time (see below sec. 6, B). The ultimate criterion of the choice of a physical theory is the agreement of theoretical predictions deduced from this theory with empirical data. However, the history of science demonstrates that this criterion usually follows the criterion of mathematical elegance. Very often, a mathematical formalism of a theory distinguishes, in some way (sometimes uniquely), the structure which leads to a good agreement with experimental results.[5]

Let us note that the analyses of the preceding section clearly suggest a correct choice of connection for the space–time model. If the new theory is to be formulated within the framework of the "frame bundle over space–time" scheme, and if it is expected to have special relativity as its limiting case, the pseudoorthogonal group $\mathcal{O}(4,1)$, called also the *Lorentz group* $\mathcal{L}(4)$, ought to be its structural group. However, according to theorem 4.4, there exist one and only one linear connection in the frame bundle which is reduced to the bundle of orthonormal frames, namely the Lorentz connection for which $\mathcal{D}g = 0$ and $\Theta = 0$. Since such a choice of connection is uniquely determined by the metric g, one can continue to denote the space–time model by the pair $(\mathcal{M}, g)$. (One should note that, according to theorem 4.4, a connection in the bundle of orthonormal frames $\mathcal{O}_1(\mathcal{M})$ uniquely determines the linear connection in the frame bundle $\mathcal{F}(\mathcal{M})$, but the reduction of the connection in $\mathcal{F}(\mathcal{M})$ to $\mathcal{O}_1(\mathcal{M})$ uniquely determines the space–time model.)

It turns out that the above criterion leads to a physical theory which remains in very good agreement with experimental data. The choice of the Lorentz connection is characteristic of general relativity, and, as is well known, this theory has successfully passed all empirical tests.

---

the reduction of the frame bundle $\mathcal{F}(X^n)$ to the bundle of orthonormal frames $\mathcal{O}_k(X^n)$ (see Sulanke and Wintgen, 1972, chapt. 2, sec. 11).

[4]There are more analogies between Lorentz metrics and Higgs fields (see Trautman 1981).

[5]Collins (1977) calls it a "subliminal role of mathematics" in contradistinction to its "sublime role". The latter consists in that "mathematics in physics acts as a language and a logical framework for discussion". Collins notices a strong increase of the "subliminal role" of mathematics in the contemporary physics of gravity.

The only degree of freedom, which is left after the above analysis, is the choice of field equations for the new theory. Einstein, motivated by reasons of a philosophical character, believed that, on the one hand, geometry of space–time should be determined by the mass–energy distribution and, on the other, motions of mass–energy should be determined by the space–time geometry. These intuitions — together with some physical arguments, such as the existence of special–relativistic and Newtonian limits — led Einstein to field equations which will now be, almost axiomatically, introduced.

Coefficients of the Lorentz connection assume the form of Christoffel symbols, and components of the curvature tensor of this connection have the following symmetries,

$$\begin{aligned} &R_{\alpha}{}^{\beta}{}_{\mu\nu} = -R_{\alpha}{}^{\beta}{}_{\nu\mu}, \qquad && R_{\alpha\beta\mu\nu} = -R_{\beta\alpha\mu\nu}, \\ &R_{\alpha\beta\mu\nu} = R_{\mu\nu\alpha\beta}, && \sum_{(\alpha\mu\nu)} R_{\alpha}{}^{\beta}{}_{\mu\nu} = 0. \end{aligned} \tag{5.1}$$

The so–called *Bianchi identities* are also satisfied:

$$\sum_{(\gamma\mu\nu)} \nabla_{\gamma} R_{\alpha}{}^{\beta}{}_{\mu\nu} = 0.$$

Contraction of the curvature tensor $R_{\alpha}{}^{\beta}{}_{\mu\beta} = R_{\alpha\mu}$ is called the *Ricci tensor*. Its subsequent (metric) contraction, $g^{\alpha\beta}R_{\alpha\beta} = R$, gives the so–called *scalar of Riemann curvature*. The Bianchi identities, after their contraction (in β and γ) and multiplication by $g^{\alpha\gamma}$, can be reduced to the form

$$\nabla_{\beta}(R^{\beta}{}_{\mu} - \frac{1}{2}R\delta^{\beta}_{\mu}) = 0. \tag{5.3}$$

Terms in parenthesis are called the *Einstein tensor* and is denoted by G. By playing with indices, one can also write

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu}. \tag{5.4}$$

It can be demonstrated that the tensor

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu}, \tag{5.5}$$

where $\Lambda = \text{const}$, is the most general two–index tensor, constructed out of the metric tensor and its first and second derivatives and having the property (5.3) (vanishing divergence). Tensor (5.5) is a purely geometric quantity. According to Einstein's idea, it should be connected with a term describing the "mass–energy distribution". A natural candidate seems to be the *energy–momentum tensor* $T_{\mu\nu}$ appearing already in special relativity. Every particle in space–time carries an energy and momentum or, more geometrically, to every non–spacelike curve in space–time at each of its points there is a tangent four–vector P^{μ}, called the *energy–momentum four–vector*. For a

particle moving with a four-velocity V^μ (with respect to a given reference frame), the energy-momentum four-vector can be obtained from the energy-momentum tensor with the help of the formula

$$P^\mu = T^\mu{}_\nu V^\nu. \tag{5.6}$$

The local conservation law can be written in the form of the vanishing divergence of the energy-momentum tensor, $\nabla_\mu T^\mu{}_\nu = 0$, which suggests the following form for the field equations

$$G_{\mu\nu} = -\kappa T_{\mu\nu}. \tag{5.7}$$

or

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = -\kappa T_{\mu\nu}, \tag{5.8}$$

where $\kappa = 8\pi G/c^4$, G being the Newtonian constant of gravity, is called *Einstein's constant of gravity*, and Λ, the *cosmological constant*. Both equations (5.7) and (5.8) were considered by Einstein as the field equations of his theory; the generalization (5.8) appeared in the cosmological context. In traditional non-cosmological applications, effects of the constant Λ are negligible and one can successfully use equations (5.7). Since Einstein the significance of the constant Λ for cosmology was widely discussed. For the sake of generality, and also because of the role played by the cosmological constant in contemporary theories of grand unification, we will usually consider the field equation in the form (5.8).

In the "empty" space-time, where $T_{\mu\nu} = 0$, equations (5.7) and (5.8) assume the form

$$R_{\mu\nu} = 0, \tag{5.9}$$

$$R_{\mu\nu} = \Lambda g_{\mu\nu}. \tag{5.10}$$

For dimensions $n = 2$ and $n = 3$, equation (5.9) implies vanishing of the curvature tensor; for dimension $n = 4$, the Ricci tensor has 10 independent components whereas the curvature tensor has 20 independent components, and consequently equations (5.9) (and of course equations (5.10) as well) have non-trivial solutions.

The field equations can be also obtained from the last action principle by assuming a suitable Lagrangian for gravity and source-fields (see, for example, Landau and Lifszic 1980). However, one should remember that, irrespectively of any theoretical reasons, the only lawful argument on behalf of any field equations is their agreement with actual empirical results. So far equations (5.7) (or (5.8)) successfully satisfy this criterion.

6. Comments and Remarks

A. A geometric space-time structure, indispensable for building up the general theory of relativity, has been introduced in terms of fibre bundle of linear frames over space time reducible to the bundle of pseudoorthogonal frames with the (homogeneous)

Lorentz group as its structural group. This is not the only possible choice. Equally natural structure would be the fibre bundle of affine frames $\mathcal{A}(\mathcal{M})$ over space–time $\mathcal{M}$ with the affine group $\mathcal{GA}(\mathbf{R}^4)$ as its structural group, reducible to the Poincaré group (see chapt. 2, sec. 3).

The fibre bundle of affine frames $\mathcal{A}(\mathcal{M})$ can be thought of as a Cartesian product of the fibre bundle of linear frames $\mathcal{L}(\mathcal{M})$ and the tangent bundle $\mathcal{T}(\mathcal{M})$, $\mathcal{A}(\mathcal{M}) = \mathcal{L}(\mathcal{M}) \times \mathcal{T}(\mathcal{M})$, with the affine group $\mathcal{GA}(\mathbf{R}^4) = \mathcal{GL}(\mathbf{R}^4) \times \mathbf{R}^4$ as its structural group, where the group $\mathcal{GA}(\mathbf{R}^4)$ acts in the following way: if $e: \mathbf{R}^4 \to \mathcal{T}_x(\mathcal{M})$ is a linear frame at $x \in \mathcal{M}$, and $u \in \mathcal{T}_x(\mathcal{M})$, $g \in \mathcal{GL}(\mathbf{R}^4)$, $b \in \mathbf{R}^4$, then

$$\mathcal{A}(\mathcal{M}) \ni (e, u)(g, b) = (eg, u + e(b)).$$

The construction of general relativity in terms of the fibre bundle of affine frames does not differ essentially from that in terms of the bundle of linear frames (see Trautman 1980, 1981). One could also choose other principal fibre bundles over space–time which would equivalently lead to the Einstein theory of gravity (see Trautman 1976).

As the analyses of the present chapter have shown, general relativity can be treated as a gauge theory with a Lorentz metric on space–time playing a role similar to that of Higgs fields, namely, spontaneously breaking symmetries to a subgroup of the structural group (see chapter 3, sec. 6). Essential differences between Einstein's theory of gravity and other gauge theories come from the existence of the soldering form in the frame bundle. This causes the fact that the geometry of the frame bundle is entirely determined by that of the base manifold, namely of space–time. (For more detail on general relativity as a gauge theory see Trautman 1978, 1980, 1981, 1984).

In sec. 5 of the present chapter, the field equations of general relativity were introduced directly as certain postulated relationships of tensorial type on space–time. By consequently using the theory of frame bundles over space–time, the fields equations can be introduced in the following way.

Let us consider the bundle of orthonormal frames $\mathcal{O}_1(\mathcal{M})$ over space–time $\mathcal{M}$, with the homogeneous Lorentz group $\mathcal{L}$ as a structural group. Let ω and Ω be connection and curvature forms on $\mathcal{O}_1(\mathcal{M})$, correspondingly. Let further g be a metric on $\mathcal{M}$, and h a biinvariant metric on the group $\mathcal{L}$ (i.e., both right– and left–invariant metrics, on $\mathcal{L}$). A metric on $\mathcal{O}_1(\mathcal{M})$ can be constructed in the following way,

$$\gamma(X, X) = g(d\pi(X), d\pi(X)) + h(\omega(X), \omega(X)),$$

where X is a tangent vector to $\mathcal{O}_1(\mathcal{M})$. This metric is invariant with respect to the action of the group $\mathcal{L}$. Let us assume $*\Omega^i_j \wedge \Omega^i_j$ to be a Lagrangian for gravity, compute the Ricci scalar R of the metric γ, and perform the variation of the action $[\int R \times$ volume element on $\mathcal{O}_1(\mathcal{M})]$ with respect to g and ω. We shall obtain the vacuum field equations with the cosmological constant together with suitable gauge equations, $D * \Omega = 0$ (see chapt. 3, sec. 6, and more extensively Trautman 1984, in particular pp. 105–107).

B. Analyses of the present chapter inevitably lead to the question: can we obtain a theory of gravity different from general relativity by choosing different connections in a principal bundle or by performing a reduction of the frame bundle to a different subbundle than that of pseudoorthogonal frames? The answer is positive. The problem has been extensively studied by Yasskin (1979). In such theories, called *metric connection theories of gravity*, the metric g, the connection ω, and the soldering form θ are dynamic variables. Dynamical equations of a given theory are obtained from the stationarity of the action $S(g, \omega, \theta, \psi)$, where ψ is a cross–section of an associated bundle which represents matter fields (see, chapt. 3, sec. 6).

The so-called *theories with a Cartan connection*, i.e., with a connection consistent with the metric but having a non–vanishing torsion, constitute an important subclass of metric–connection theories of gravity. The *Einstein–Cartan theory*, belonging to this subclass, was especially thoroughfully investigated. The torsion in this theory is interpreted as macroscopic spin of particles. Studies concerning this theory go back to Weyl (1919) and Cartan (1922, 1923, 1924, 1925); afterwards the theory was developed by Kibble (1961); Sciama (1962); Hehl (1973, 1974); Trautman (1972 a–c, 1973 a–c, 1975, 1978, 1980, 1981).

Bibliographical Note. More about the structure of the frame bundle can be found in the literature dealing with bundles in general (see *bibliographical note* to chapter 3), in the works by Trautman (1970b, 1976), and especially in the review paper by Dodson (1979).

The presentation of general relativity in sec. 5 of the present chapter was very concise. The reader should consult at least some expositions of the huge literature concerning this beautiful theory. I shall quote only some examples.

The following textbooks are often recommended: Weinberg (1972), Misner et al. (1977), Landau and Liphshitz (1980), Rindler (1977), Schutz (1983). To classical textbooks, which always deserve attention, belong Weyl (1922), Eddington (1923), Tolman (1934), Robertson and Noonan (1968).

Mathematical structure of the theory was elaborated in the following monographs: Petrov (1966), Hawking and Ellis (1973), Sachs and Wu (1977), Carmeli (1977), Beem and Ehrlich (1981). O'Neill (1983) can serve as a good introduction to the mathematical structure of general relativity. No serious work concerning solutions to Einstein's field equations can be done without consulting the fundamental monograph by Kramer et al. (1980). A useful collection of exercises is in Lightman et al. (1975).

Philosophical aspects of general relativity are explored by Tonnelat (1971), Raine and Heller (1981), Friedman (1983), and Toretti (1983).

Chapter 5

The Cosmological Problem

0. Introduction

Relativistic cosmology, or more precisely its theoretical foundation, is a main topic of the present book. The foundations of cosmology are deeply rooted in the contemporary theory of space–time and gravitation. Therefore, the preceding chapters constitute a preparation for the main topic, or perhaps, even more than a preparation: as I shall try to demonstrate in the present chapter, global aspects of space–time and gravitation belong, in a sense, to the field of cosmology. Now, the cosmological problem will be addressed explicitly.

Cosmology aims at investigating the structure of space–time on its largest scale. One often speaks of the Universe, having in mind not only the space–time itself, but also everything that fills it in, i.e., all physical fields defined on space–time. All these concepts are understood more or less in an intuitive way. However, one must admit that it is not a rough intuition, but an intuition shaped by long tradition, especially by the evolution of concepts, which has acquired a great acceleration in our century.

There exist two essentially different "directions" of studying cosmology. The "down–direction" consists in assuming some *a priori* postulates concerning the world structure at large and trying to deduce the "local physics" from them. The "up–direction" accepts the local physics as given and tries to extrapolate from it as far as possible to guess at the global structure of space–time. The first way of doing cosmology was represented by Milne (1935, 1948), and Bondi and Gold (1948); but it was the second approach that has acquired the paradigmatic citizenship in the cosmology of the XX^{th} century. The extrapolating cosmology also cannot avoid doing some more or less aprioristic assumptions concerning either the Universe itself or the nature of cosmology as a science of the Universe. Extrapolation, from its very essence, goes beyond local collecting of information, and if it does not explicitly make some assumptions supervising this process, this means it has made such assumptions implicitly. Unfortunately, the majority of publications in cosmology (not to say all with very few exceptions) choose the second possibility. This is an additional motive to write a chapter in a book on foundations of cosmology discussing such problems, even at the risk of loosing something of the formal consistency of the book. At the present state of its development the analysis of the cosmological problem cannot be fully mathematical. However, an analysis seems indispensable for a mathematical

formalism to be able to correctly fill in with a cosmological content. This justifies a different, rather unformal, character of the present chapter.

In sec. 1, I try to determine what it means that cosmology is a global geometry of space-time or a non-local physics. Quite unexpectedly, it turns out that non-local assumptions (with a natural understanding of "non-local") are involved in every empirical prediction of standard physics. Global character of cosmology can be understood in a strong sense as a study of the set of all solutions to cosmological equations (Einstein's field equations), i.e., of the set of all cosmological models (such a set is called an *ensemble of universes*). Understanding of cosmology as a theory of such a set is not only a useful methodological trick but it has been implicitly assumed in the scientific practice for a long time. These problems are considered in sec. 2. One cannot forget that cosmology has the ambition to belong to the family of empirical sciences. Consequently, the cosmological problem cannot be responsibly dealt with without touching its observational aspects. These aspects certainly deserve a separate monograph. One can hope that the works of Ellis' group (quoted in the present chapter) will lead to such a monographic study. Therefore, in sec. 3, I only summarize the most important results of the Ellis programme. These results strengthen the above formulated thesis that the discussion of empirically unverifiable assumptions, which enable cosmology to be a science of non-local structure of space-time, must enter the core of the cosmological problem.

1. Cosmology as a Global Geometry of Space-Time

As we know from the preceding chapters, all dynamical theories of hitherto physics presuppose the manifold structure of space-time. However, the traditional approach to physics makes use almost exclusively of local properties of the manifold model. What matters is only a "correct behaviour" in the neighbourhood of the point at which experiments or observations have been made. In practice, one is interested in how "to translate statements about geometrical entities in space-time into statement about real numbers" (Friedman 1983, p. 33), and the work is smoothly done by local maps which define the manifold structure of space-time. On the other hand, the very concept of a manifold is a non-local notion, and the assumption that physics can develop by using only local methods, i.e., that a "local neighbourhood" could be isolated from global influences, turns out to be a strong assumption of a non local character.

From the field-theoretical point of view the situation is the following. Local empirical predictions are trustworthy only if unexpected perturbations are excluded in the future comprised by the prediction. In order to predict the state of a field at a certain instant one must know "initial conditions" at an "earlier" spacelike hypersurface (or at a null cone, see Penrose 1980). The assumption that there are no signals, generated by a distant event, which at the next moment could affect the investigated system ("no-news" assumption, see Ellis 1983), implies a knowledge about the field

state on a distant (towards the past) hypersurface. In this sense, cosmological assumptions are contained in every empirical prediction (see Heckmann and Schücking 1959).

The above considerations justify the following assertion: *The Universe possesses a certain structure (or it is a structure); this structure can be investigated by using either local or non–local (global) methods; the first is a task of physics, the second — of cosmology.* For the time being, terms "non–local" and "global" are used interchangeably. In the following, I shall explain the meaning of the above thesis and argue on behalf of it. Let us now pause for a while to reflect upon a difference between local and global methods. Even in contemporary mathematics, a substantial part of which consist of "global analysis", these terms are fuzzy and non–precise. In Smale's opinion, "global analysis is simply the study of differential equations, both ordinary and partial, on manifolds and vector space bundles" (Smale 1980, p. 84). As we can see, more or less, such a definition of "globality" has been accepted in the preceding chapters. The manifold model of space–time and fibre bundles over space–time were for us fundamental research tools from the very beginning.[1]

Modern observational astronomy deals with such distant objects (galaxies, radiogalaxies, quasars, ...) that no reasonable interpretation of what is observed can be made without a theory of the space–time over which received signals are propagated. This means that some features of the Universe, investigated by contemporary astronomy, essentially depend on integration of some differential equations over a "large scale" domain of the space–time manifold. In this sense cosmological methods permeate the domain of the extragalactical astronomy. Very often this is obscured by the fact that astronomers usually employ the simplest cosmological models ascribing to the Universe strong symmetries (such as Robertson–Walker symmetries). Such symmetries effectively eliminate the necessity to compute troublesome integrals over space–time. Of course, this fact substantially simplifies the situation but, on the other hand, it "screens out" the role of cosmology in standard astronomical investigations (see Ellis and Sciama 1972).

General relativity suffers a certain dualism. Space–time manifold, equipped with different geometric structures, is a "primitive element" of the theory, but on this space–time manifold various physical fields are defined. It is true that these physical fields are also represented by certain geometric structures on space–time (such as scalar or tensor fields); however, they do not appear naturally in the process of the development of the theory, but are introduced into it by decree. Einstein's field

[1] A tradition of global analyses in mathematics goes back to Poincaré. To denote this kind of research, Morse in 1934 used the term "macroanalysis". He noticed that "any problem which is nonlinear in character, which involves more than one coordinate system or more than one variable, or whose structure is initially defined in the large, is likely to require considerations of topology and group theory in order to arrive at its meaning and solution" (quoted after Smale 1980, p. 118). The famous Morse theory can serve as an example of non–local methods in mathematics. This theory, to a large extent, contributed to developments of the global analysis. (See, Smale's essay *What Is Global Analysis?* in Smale 1980, pp. 84–89; see also ibid. pp. 117–127).

equations (see chapt. 4, sec. 5) relate geometry and physical fields with each other. In this way, a non-local character of geometrical methods in cosmology implies a non-local character of physical fields (with the exception of gravitational field which is already incorporated into geometry). One could simply say that cosmology is a non-local physics.

It is interesting to notice that in such exotic (from our present point of view) states at the first fractions of a second after the initial singularity, the very distinction between local and non-local ceases to be clear cut. The closer we are to the initial singularity, the more strongly will any physical property imply non-local consequences.[2]

The initial singularity itself turns out to be a non-local property of the model. From the geometric point of view singularities are understood as geodesic incompleteness of space-time. A space-time is geodesically incomplete if there is at least one non-spacelike geodesic in it along which the affine parameter cannot assume arbitrarily large values (see Hawking and Ellis 1973). The well-known Penrose-Hawking Geroch singularity theorems (see ibid.) show that if a space-time possesses certain properties, they cannot be "fitted together" with the requirement of the geodesic completeness. Among those properties there are always some of a manifestly non-local character, for instance: spatial compactness, non-existence of closed timelike curves, existence of a Cauchy surface, etc. It is worthwhile to note that in the initial singularity of the standard cosmological model (in the so-called *Big Bang*) all non-spacelike geodesics break down; this fact additionally strengthens a global character of the singularity.

Even in space-times in which not all non-spacelike geodesics break down at a singularity, the appearance of a singularity causes the appearance of horizons and boundaries which are typically global properties of space-time (see Penrose 1968, Hawking and Ellis 1973, Geroch and Horowitz 1979, Tipler et al. 1980).

There are different kinds of singularity (see Clarke 1975, Clarke and Schmidt 1977, Ellis and Schmidt 1977, Tipler et al. 1980), and some of them are "more global than others". A singularity is said to be a *curvature singularity* if in a given space-time there exists at least one non-spacelike curve such that the components of the curvature tensor, in a frame parallelly propagated along this curve, diverge as one approaches the singularity. A singularity is said to be *quasi-regular* if in such a situation components of the curvature tensor, in any such frame, behave correctly. Suppose that an observer approaches a quasi regular singularity. He is in no way locally warned of the imminent danger. Indeed, on the strength of the Clarke theorem (1973), his history (a timelike curve) has a neighbourhood isometric to a one in a complete space-time manifold. "This means that the 'bad behaviour' that causes the incompleteness is of a global nature: near a single curve, one does not realize that anything is wrong" (Tipler et al. 1980, p. 159). Ellis and Schmidt (1977) have shown

[2]This idea is contained in the primeval atom hypothesis of Georges Lemaitre (see Godart and Heller 1985).

that space–time in a neighbourhood of a quasi–regular singularity can be thought of as being glued together out of several copies which are perfectly regular. As we can see, in such a situation everything seems to be in order, and the geodesic incompleteness has clearly a global character.

2. Cosmology as a Theory of the Ensemble of Universes

In mathematics, there is also another understanding of what global could mean. We face it when, for example, we are dealing, not with a single differential equation on a manifold $\mathcal{M}$, but with a space of all possible equations (of a given type) on a manifold $\mathcal{M}$. Steve Smale, in his short essay entitled *What is global analysis?*, after having quoted Peixoto's theorem asserting that the structurally stable differential equations on a compact two–dimensional manifold $\mathcal{M}$ form an open and dense set, writes: "This theorem is an excellent theorem in global analysis. One sees in two ways how it is global. First, the differential equation is defined over a whole manifold, and structural stability depends on its behaviour everywhere. Second, the theorem makes a conclusion about the space of all differential equations on $\mathcal{M}$" (Smale 1980, p. 88). In the preceding section we have explored the first meaning of the globality as applied to cosmology, now we shall comment upon its second understanding (which could be termed "superglobality"). In general relativity things are much more difficult than in any "classical" theory of differential equations. Einstein's equations themselves determine the manifold on which they act, and Smale's second understanding of what global means leads to much more difficult — and, one must admit, philosophically much more fascinating — problems. There is no other way; such an understanding must be reduced to the study of all admissible solutions to Einstein's field equations. A mere look upon the history of the twentieth century cosmology reveals that it has developed, as far as its theoretical aspects are concerned, by constructing newer and newer cosmological models, that is to say, by exploring the space of all possible solutions of Einstein's equations.

In agreement with a commonly accepted custom among relativists, the space of all solutions of Einstein's field equations will be called the *ensemble of universes* or shortly the *ensemble*. We shall try to look at cosmology as at a theory of the ensemble. We will argue that such an approach to cosmology is unavoidable.

Here we meet a crucial point. It is commonplace, both in scientific and in popular literature, to stress that all peculiarities of cosmology, as a science, follow from the fact that "the Universe is given to us in one copy". I know of no research work that would go more deeply into this methodological peculiarity of the science of the Universe. I will dwell on the problem for a while.

For the empirical method it seems essential to have many instances of a similar type. Laws of physics are formulated in the form of differential equations which express a structure built up from relations among quantitative properties common to many phenomena. Individual characteristic of every phenomenon are accounted for

by choosing suitable initial or boundary conditions which select a particular solution from the class of all admissible solutions. The uniqueness of the Universe makes it impossible to transfer this strategy directly to cosmology. However, one could hardly see how to develop a science otherwise. Cosmology, therefore, has chosen the following tactics. One assumes that a structure called the Universe is described by a solution of a certain equation (usually, Einstein's field equations); one attempts to find all its possible solutions, and then to identify those initial or boundary conditions which correspond to the Universe we actually observe. All other solutions describe other possible universes. Owing to this strategy the actual Universe is placed in the environment of many universes. Now, one has many "individuals", and one can proceed in the same manner as in other departments of physics. In this sense, as Bondi (1960, p. 10) remarks, "we select the important (as opposed to the 'accidental') features of the actual universe by their relation to the theory chosen rather than by any independent criterion". In the following, the set of all possible solutions to Einstein's field equations will be called the *ensemble of universes*, or the *ensemble*, for brevity.

The richness of all possible solutions of Einstein's field equations is so great that usually we explore only "small" subsets of the set of all possible solutions (for instance, spatially homogeneous or homogeneous and isotropic solutions). However, the fact that we can explore subsets of the ensemble is essential for cosmology. Recently, one can notice quite a pronounced tendency: it seems that cosmologists are more interested in, say, how a given property (e.g., the existence of singularities, the appearance of horizons, etc.) is distributed within the ensemble rather than in the structure of individual world models (of single points of the ensemble). This tendency could be explained by methodological reasons. Because of the measurement error inherent in every cosmological observation, a solution which pretends to describe the real Universe can do it only up to a finite degree of accuracy. If we improve our measurements, we must be ready to change to some "nearby" solution which could describe the Universe to a better approximation. In this sense, a *structural stability* of solutions must be postulated (see Thom 1977, Arnold 1980). The problem is by no means a trivial one. It must be decided which solutions are to be defined as identical (or equivalent), and when the distance between two different (non-equivalent) solutions is to be said "small". The decision must be both mathematically and operationally realistic. From the mathematical point of view, the space of all solutions must be equipped with a metric, or at least with a suitable topology. From the operational point of view, the adopted definitions should correspond to what, and with which accuracy, is measured in observational cosmology.

The stability property is also interesting with respect to other features, not necessarily directly observable, which we would like to ascribe to the actual Universe (for instance with respect to such properties as the existence of singularities or horizons, different degrees of causality, and so on). If a property of a certain model is shared by "neighbouring" models it is said to be a *typical property* (or a *generic property*).[3]

[3]More precisely: a subset of a topological space is said to be a *Baire set* if it is an intersection

This presupposes a metric or at least a topology on the ensemble which have to be both mathematically workable and leading to physically reasonable results. So far a "natural" topology on the ensemble is not known, and to investigate the stability of different properties one must use different topologies (see below). Theoretical aspects of the structural stability of certain classes of cosmological models were investigated, among others, by Collins (1974, 1977), Golda et al. (1983), Szydłowski et al. (1984); observational aspects of structural stability still await elaboration (see remarks at the end of sec. 3).

The notion of the ensemble of universes was implicitly used in cosmology for a long time. For instance, the Cauchy problem, when considered within the cosmological context, leads to this notion almost immediately: the set of all admissible Cauchy data (on a certain spacelike hypersurface) may, quite naturally, be identified with the set of all universes which can be developed out of these data. Explicitly, the concept of the ensemble began to emerge in the work by Dicke (1961) who, for rather philosophical reasons, postulated that there exists not only the Universe but a whole class of universes with all possible values of physical constants. These ideas found a more elaborate form in Carter's talk during the Symposium of the International Astronomical Union in Cracow, 1973. Carter used the concept of the ensemble of universes to formulate a doctrine called by him the anthropic principle. He used also the name "world–ensemble". "By this I mean — he wrote — an ensemble of universes characterized by all conceivable combinations of initial conditions and fundamental constants (the distinction between these concepts, which is not clear–cut, being that the former refers essentially to local and the latter to global features)" (Carter 1974). According to the author, "the existence of any organism describable as an observer will only be possible for certain restricted combinations of the parameters, which distinguish within the world–ensemble an exceptional cognizable subset". Since we are certainly "describable as organisms", the Universe we are living in must belong to this distinguished subset.[4]

---

of countably many open and dense subsets of this space. A property is called *typical* if it is shared by all elements of a Bair set. Property of the complement of a Bair set is called *non–typical*. The difference between typical and non–typical properties cannot be reduced to whether a corresponding set is dense or not (complements of Bair sets can be dense) or to the measure criterion (in the case of the space of a finite dimension, a Bair set can have zero Lebesgue measure). However if, between two contradictory properties that can be possessed by a model, one is typical, the other must be non–typical, and vice versa. An important notion is that of a strongly typical property. A property is *strongly typical* if it is possessed by elements of an open and dense subset of a given topological space (some authors, when writing of typical properties, have in fact in mind strongly typical ones). Correspondingly, one speaks about *strongly non–typical properties*. If a model has a strongly typical property, it preserves it after being slightly perturbed. Moreover, a property is preserved if a perturbed model is perturbed again provided the second perturbation is "smaller" than the first one. On the contrary, if a model has a strongly non–typical property, a slight perturbation can always be found which liquidates this property. (See Richtmyer 1981, pp. 304–311).

[4]Later on, the anthropic principle gained considerable popularity among cosmologists (see, for instance, Carr and Rees 1979). The possibility of reintroducing the man back into the post–Copernican

In the last decade — independently of any philosophical inspirations — the space of solutions of Einstein's field equations was a subject of intensive mathematical research. Usually, one first considers the space $\mathrm{Lor}(\mathcal{M})$ of all Lorentz metrics on a manifold $\mathcal{M}$. Serious difficulties are encountered already at this stage. It turns out that one can introduce infinitely many topologies on $\mathrm{Lor}(\mathcal{M})$, and neither of them is natural. To show the nature of the problem let us consider it in a more detailed way.

The neighbourhood of a metric $g \in \mathrm{Lor}(\mathcal{M})$ can be understood as a set of Lorentz metrics on $\mathcal{M}$, whose derivatives up to an order r only slightly differ from the corresponding derivative of the metric g on a compact set $U \subset \mathcal{M}$. The topology defined with the help of such neighbourhoods is called *C^r–compact–open topology* on $\mathrm{Lor}(\mathcal{M})$. If one assumes $U = \mathcal{M}$, one obtains the so–called *F^r–open topology* on $\mathrm{Lor}(\mathcal{M})$.

It turns out that every C^{r+1}–compact–open topology is richer than C^r–compact–open topology. Similarly, every F^{r+1}–open topology is richer than F^r–open topology. And every F^r–open topology is richer than C^r– compact–open topology.

Moreover, one can define the so–called W^r *Sobolev topology* on $\mathrm{Lor}(\mathcal{M})$, by replacing the postulate of the closeness of metric derivatives with the postulate of the closeness of integrals of squares of these derivatives.[5] It can be shown that every W^{r+3} (compact–open) Sobolev topology is richer than C^r–compact–open topology which, in turn, is richer than W^r Sobolev topology.

This hierarchy of topologies on $\mathrm{Lor}(\mathcal{M})$ can be further enriched in many ways (for more details see Hawking 1971; Hawking and Ellis 1973, pp. 198, 252–253; Beem and Ehrlich 1981, pp. 28–29). Neither of these topologies is "natural" on $\mathrm{Lor}(\mathcal{M})$, and one must suitably choose different topologies for different problems. For instance, Sobolev topologies turn out to be suitable for discussing the stability of initial conditions for Einstein's equations[6], but to define the so-called stable causality[7] compact–open topologies are preferred. However, when considering the ensemble problem, we are not interested in all Lorentz metrics on $\mathcal{M}$, but in all those Lorentz metrics on $\mathcal{M}$ which are solutions of Einstein's equations; they constitute the subset $\mathcal{E}$ of the space $\mathrm{Lor}(\mathcal{M})$. The space $\mathcal{E}$ usually has the structure of a smooth manifold. Moreover, there exists its local representation in the space of four functions of three variables. It turns out that Einstein's equations act in $\mathcal{E}$ as a Hamiltonian system.

One–parameter families of solutions for Einstein's equations are curves in $\mathcal{E}$,

---

universe caused far too much of metaphysical emotions; see a responsible review by Demaret et Barbier (1981) and a balanced evaluation by Barrow (1983). However, the main reference to the anthropic principle, as far as its both physical and metaphysical aspects are concerned, is a book by Barrow and Tipler (1983).

[5]These integrals and squares are defined in terms of a certain positively defined metric on $\mathcal{M}$ which always exists.

[6]The problem is how to formalize the postulate that a slight change of the initial conditions leads to evolutions only slightly different from each other.

[7]Space–time $(\mathcal{M}, g)$ is said to be *stably causal* if a slight perturbation of the metric g does not produce closed timelike curves.

and spaces of solutions of linearised Einstein's equations form tangent spaces to $\mathcal{E}$ (fig. 5.1). Of course, one can meaningfully speak of a tangent space to $\mathcal{E}$ only if $\mathcal{E}$ has the structure of a smooth manifold in the neighbourhood of a given solution. In such a region, Einstein's equations are said to be *stable with respect to linearization*. This is not always the case. One of the most interesting results, as far as investigations of the ensemble are concerned, is the theorem stating that Einstein's equations (with the vanishing energy–momentum tensor) in the neighbourhood of a solution $g_0 \in \mathcal{E}$, such that the space–time $(\mathcal{M}, g_0)$ has a compact Cauchy surface, are stable with respect to linearization if and only if the space–time $(\mathcal{M}, g_0)$ admits no Killing vector field (see Fisher and Marsden 1979, p. 206). This result points to the fact that space–times with symmetries (with Killing vector fields) violate the smooth manifold structure of the solution space $\mathcal{E}$. Indeed, it turns out that the space $\mathcal{E}$ has a cone singularity in the naighbourhood of a solution with symmetries (see Fisher et al. 1980; Brill 1982).[8] This fact has serious consequences in the theory of "small perturbations" of Einstein equations (e.g., in the problem of the origin of galaxies and clusters of galaxies). Since in the neighbourhood of a solution with symmetries the space $\mathcal{E}$ is not stable with respect to linearization, small perturbations of such solutions produce, in principle, spurious perturbations (unless they satisfy the so–called Taub condition, see Fisher et al. 1980).

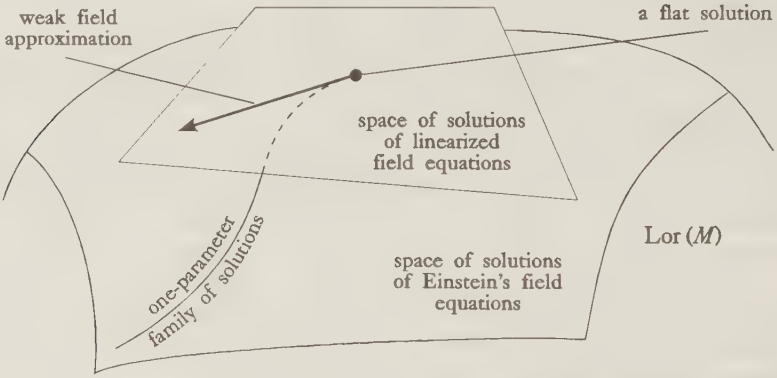


Fig. 5.1. The space of solutions of Einstein's equations.

Let these few results suffice as illustration of the subtle nature of the ensemble. Our aim is not to review this field of research (for more see Montcrief 1975; Marsden 1981, chapters 9 and 10, and already quoted works), but only to point out that the ensemble, from its very nature, is a natural environment of cosmology.

[8] Analogously to the "cone" $F(X, Y, Z) = X^2 + Y^2 - Z^2$ at the point (0, 0, 0). Everywhere besides this point, the linearized equation $dF/ds|_{s=0} = 2XX' + 2YY' - 2ZZ' = 0$ uniquely defines the tangent space.

It might turn out that the space of all solutions approach to cosmology be indispensable in formulating quantum theory of gravity and its applications to understanding the evolution of the Universe. For instance, Hawking (1979, 1982) argues that a suitable method to quantize gravitational field should consist in adapting to this goal Feynman's path–integral method. The method, as applied to gravitational field, consists in specifying a three–metric h_1 and the values of matter fields φ_1 on a spacelike surface $\mathcal{S}_1$, and going to a three–metric h_2 and the values of matter fields φ_2 on a spacelike surface $\mathcal{S}_2$, as a sum over all four–geometries having $\mathcal{S}_1$ and $\mathcal{S}_2$ as their boundaries, with matter fields suitably matching φ_1 and φ_2 correspondingly. The method was elaborated and applied to cosmology by Hartle and Hawking (1983) and by Hawking himself (1984). These authors assumed that spacelike three–geometries are compact (the closed universe). Quantum states of such a universe are described by a wave function defined as a functional on the compact three–geometries and the values of matter fields on them, and satisfying a second order functional differential equation, called the *Wheeler–DeWitt equation*. Specifying the states of the universe is equivalent to specifying the boundary conditions for this equation.

By a rotation to Euclidean time τ ($t \mapsto -i\tau$) in the action, one changes to the integration over all positive–definite four–geometries having suitable three–geometries as their boundaries (the so–called Euclidean approach, see Hawking 1979). Hartle and Hawking put forward a proposal that the ground–state amplitude for a given three–geometry is that given by an integral over all *compact positive–definite* four–geometries having this three–geometry as a boundary.[9] "This means that the Universe does not have any boundaries in space or time (at least in the Euclidean regime). One can interpret the functional integral over all compact four–geometries bounded by a given three–geometry as giving the amplitude for that three–geometry to arise from a zero three–geometry, i.e., a single point. In other words, the ground state is the amplitude of the Universe to appear from nothing" (Hartle and Hawking 1983, p. 2961).

As it can be seen, the Hartle–Hawking method presupposes the space of all possible states of the Universe (of all compact three–geometries). This set is called *superspace* (having nothing in common with the superspace of supergravity theory), and in quantum cosmology it plays a similar role to that played by the ensemble in classical cosmology.

Finally, let us note that historical developments of the relativistic cosmology can be naturally described in terms of the ensemble. Theoretical investigations in cosmology had penetrated larger and larger domains of the ensemble by discovering more and more new solutions. Later on theoreticians were interested in how certain properties of world models were distributed in the ensemble. It is interesting that world models first discovered historically (the static Einstein model, the empty de Sitter model ...) or classes of world models (spatially homogeneous but isotropic

[9] Matter fields should be regular on these geometries, and only space–times with the cosmological constant $\Lambda > 0$ are considered, $\Lambda = 0$ is regarded as a limiting case.

models, spatially homogeneous but anisotropic models ...) usually turned out to be non–typical in the space of solutions. The direction of research in observational cosmology is just the opposite: investigations in this field always tend to narrow a certain domain in the solution space, i.e., to distinguishing this class of solutions (or, in the highly idealized case, this single solution) which best fits the actual observational data. Without its observational image modern cosmology would loose its identity. Only if we remember about its observational side can we claim that cosmology is a theory of the ensemble.

3. Observational Problems of Cosmology

Since Lemaître's work of 1927, in which this author for the first time demonstrated that actual red shift measurements in galactic spectra do not contradict predictions of a relativistic world model, comparisons of theoretical models with observational data became an important chapter of cosmology. Hubble's (1936, 1937) ambition was to treat cosmology as an observational science in the following sense: the task of astronomical research is to describe the "observable universe"; if one assumes that it provides a "fair sample of the Universe", a simple extrapolation will allow one to arrive at a satisfactory cosmological model. However, soon it became evident that Hubble's approach was unrealistic. Modern astronomical (and radioastronomical) techniques reach such faraway domains of space–time that it is impossible to interpret the registered "data" without assuming a space–time model (at least as a working hypothesis). In practice, the following procedure, called the *empirical testing of cosmological models*, has been established. Basing on certain postulates, usually inspired by philosophical views (such as cosmological principle, postulate of the spatial closure of the Universe, and so on) or by empirical results believed to be well–founded (e.g., symmetrical matter distribution in the Universe), a certain cosmological model or a class of models is assumed. In the second step, one deduces relationships between observable magnitudes (the so–called *observables*) from this world model or from this class of world models, and then one compares theoretically deduced relationships with those established observationally. Results of this procedure were initially only able to show that there were no contradictions between a given model (or a given class of models) and the observational data.

The first major success of the observational cosmology consisted in making the steady–state cosmology (Bondi and Gold 1948, Hoyle 1948) "non–satisfactory" and "less attractive" than its rival evolutionary world models. However, this became possible, not owing to the traditional method of observational testing, but rather to the discovery of the microwave background radiation (Penzias and Wilson 1965, Dicke et al. 1965). The steady–state cosmology was not able to explain the genesis of this radiation.

In the seventies the so–called *standard cosmological model* has been established as a result of several factors. Accumulation of new empirical data indeed seemed to

corroborate postulates assumed so far on purely theoretical basis (see, for instance, Symposium of the International Astronomical Union in 1973 on comparing observational data with theoretical models (Longair 1974)). An important role was also played by stronger and stronger connections between the evolutionary world model and standard physics: high energy physics has found its "natural laboratory" in the very early Universe. On the other hand, typically relativistic methods began to infiltrate the field of astrophysics. People get gradually accustomed to the fact that the Universe is a relativistic object. However, during this entire period the method of testing of cosmological models was based on intuitions and many tacit assumptions, believed to be self-evident, rather than on a solid mathematical analysis of the problem. It is symptomatic that a paper by Sandage (1961) was long treated as a classical work in this field in spite of the fact that it dealt with observational possibilities of the 200-inch telescope rather than with observational foundations of cosmology.

A paper by Kristian and Sachs (1966) was the first in which the authors began systematically to analyse the problem. They treated it in a general way, not limiting themselves to a chosen class of models. They assumed that (1) space–time is a four-dimensional pseudo–Riemannian manifold with a structure enabling one to express all the investigated observational relationships as a series around the here–now of the earthly observer; (2) null geodesics are histories of light beams; (3) the dynamics of the world is ruled by Einstein's field equations with the dust energy–momentum tensor. Assumption (1) is the most limiting one. It implies that all investigated relationships remain correct for space-time distances which are small when compared with both the scale factor of a given world model and the scale of local inhomogeneities. In the case of relativistic cosmology, assumption (2) can be deduced from the postulates of general relativity as the limit to geometric optics. In the work of Kristian and Sachs, assumption (3) interferes only with calculations leading to numerical estimates of some observational magnitudes; the authors succeeded in making no use of dynamical equations in all other considerations.

However, cosmology owes systematical analysis of its observational possibilities to G.F.R. Ellis and his group. In a series of works (Ellis 1975, 1980, 1984; Ellis and Perry 1979; Ellis et al. 1978) the problem has been clearly stated, and some of its aspects discussed in a detailed way. An extensive paper of Ellis et al. (1985) marks the close of a stage of program implementation. It certainly deserves a closer look.

In their first work, the authors attempted at decoding information of cosmological significance contained in the ideal astronomical observations, i.e., with the assumption that all measurements can be performed with any desired precision and without making use of any dynamical equations. The assumption remains in power that the pair $(\mathcal{M}, g)$, where $\mathcal{M}$ is a four-dimensional smooth manifold, and g a Lorentz metric defined on it, is a model of space–time. As usual, null geodesics are histories of photons, and timelike geodesics histories of freely falling particles with non-zero rest-mass. The program consists in expressing the metric g with the help of the so-called *observational coordinates*, i.e., such coordinates as can be directly

translated into observational terms. If derivatives of the components of the metric tensor (in observational coordinates) can also be determined by measurements, then, by computing the coordinates of the connection and of the curvature tensor in the usual way, one could fully reconstruct the space–time model from astronomical observations. In this method one need not expand any observable into a series, and consequently one need no assumption about the smallness of the region controlled by actual observations.

The observational coordinates $\{x^i\} = \{w, y, \theta, \varphi\}$ are defined with respect to the history C of our Galaxy or of our Cluster of Galaxies; one assumes that this history is a suitably smooth timelike geodesic (at least in a neighbourhood of the instant in which observations are made) (see fig. 5.2). The surfaces $w \equiv w_0 = \text{const}$ are past light cones of events situated on the curve C. The parametrization of the curve C is normalised in such a way that w measures the proper time along C. The instant $w = w_0$ corresponds to the event q which is our "here and now".

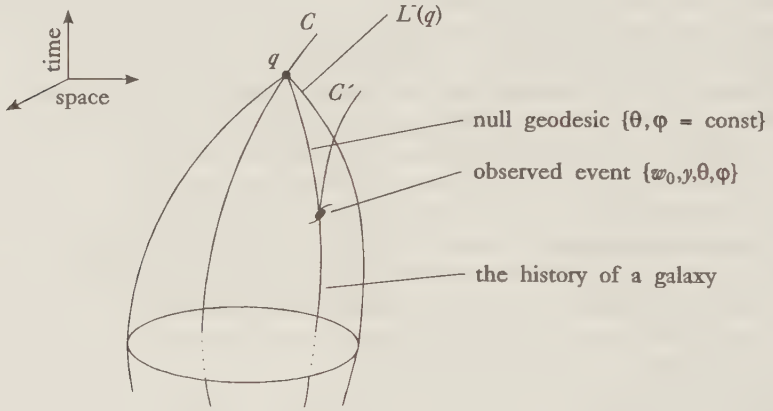


Fig. 5.2. Observational coordinates.

Let us denote the light cone of the event q by $L^-(q)$. Null geodesics generating $L^-(q)$ are given by $\{x^2 \equiv \theta = \text{const},\ x^3 \equiv \varphi = \text{const}\}$ on the surface $w = \text{const}$. The coordinates θ and φ can be normalised in such a way that they become the usual spherical coordinates on the celestial sphere with respect to a non–rotating reference frame.

The coordinate $x^1 \equiv y$ measures space–time distance along null geodesics. One should notice that the concept of a space–time distance contains both a space distance to the curve C and a temporal distance to the event q. There are a few possibilities as to how practically identify the coordinate y with the actually measured "distances".

It may happen that the coordinates $\{w, y, \theta, \varphi\}$ do not cover the entire space–time manifold (even if one does not take into account the coordinate singularity

$\theta = \varphi = 0$). However, they do cover the region $\mathcal{N} \subset \mathcal{M}$ which can be observed from the curve C (these coordinates can also give a "one-many" covering, i.e., the same coordinates can designate distinct events; in such a case, ghost images of galaxies may appear).

Usually, one assumes that the length of the period during which the humanity performs its astronomical observations is negligibly small as compared with the age of the Universe, and one can treat it as an event q on the curve C. If — with a bit of optimism — one replaces the event q by a finite proper time interval $\mathcal{I}$, the observational situation is slightly improved (fig. 5.3).

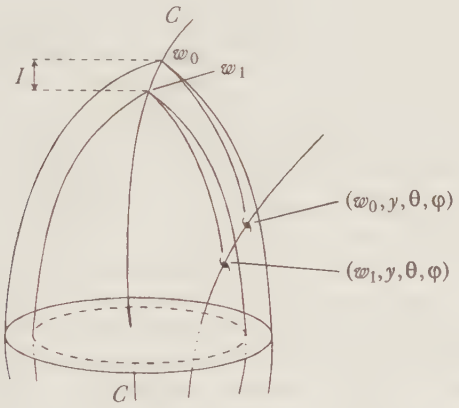


Fig. 5.3. Observational situation with the assumption of a finite period of performing observations.

The general conclusion from the detailed analysis — which, unfortunately, cannot be carried out here — is rather pessimistic (as it should be expected): ideal astronomical observations are not enough to uniquely reconstruct the geometry of space–time on our past light cone $L^-(q)$. It turns out, in particular, that we are not able to observationally determine (with no help of dynamical equations) how the three–dimensional hypersurface forming our past light cone is embedded in the four–dimensional space–time manifold. Moreover, it is impossible to demonstrate observationally that space–time is spherically symmetric around us; neither the opposite statement can be demonstrated. These pessimistic conclusions remain in force even if one assumes that our observations cover a finite span of time. Therefore, observational data collected from $L^-(q)$ leave a considerable cosmological indeterminacy: without help from dynamical equations (such as Einstein's field equations) many different world models can still show consistency with observational data.

How the situation is improved when one takes into account dynamical equations? One should be aware of the fact that after such equations have been accepted, obser-

vational data become theory–impregnated to a much higher degree than before; and one cannot hope that one would still be able to discern in a cosmological model what comes from observation and what is theoretically introduced into it. The problem was addressed in the second part of the work by Ellis et al. (1985). Einstein's field equations have been assumed as dynamical equations in cosmology; therefore, in this work, observational implications of the relativistic cosmology are investigated. No space–time symmetries are *a priori* assumed. On the contrary, the aim of the work is to demonstrate the existence or non–existence of space–time symmetries (or of other properties of a model) from observations helped by the theory of general relativity.

The observational coordinates are also used in this approach, but they have to be adapted to the field equations (afterwards they were analysed with the help of the Newman–Penrose formalism). The problem consists in determining (in terms of ideal astronomical observations) the Cauchy data for the field equations on the past light cone $L^-(q)$, and then in integrating these equations on and outside the past light cone $L^-(q)$. However, one must formulate the *final value problem*, characteristic for cosmology: one integrates the equations from the Cauchy data on $L^-(q)$ backward to a certain value $y = y^*$. This limitation for the value of y is a consequence of the fact that the very early Universe was not transparent to electromagnetic radiation. The backward integration on the light cone $L^-(q)$ can meet obstacles in a caustic form which may have origin either in the topology of space–time or in the phenomenon of *gravitational lensing*. This, of course, creates serious difficulties for the program of observational cosmology (see Friedrich and Steward 1982).

By using the field equations, the coordinate y can be (uniquely) replaced by the red shift z. To the limiting value y^* now corresponds the value z^*.

The Cauchy data, consisting of the most detailed cosmological information we could hope to obtain, form the *maximal data set*; let it be denoted by $\mathcal{D}(w_0, z)$. It consists in the following ideal observations: red shifts, proper motions, distances to galaxies (the so–called area distances)[10], distortion of images of celestial objects and their groups (provided that "proper shapes" are known), results of galaxy and quasar counts up to a limiting value $z = z^*$. It is worthwhile to notice that the background radiation is not an element of $\mathcal{D}(w_0, z^*)$. This radiation does not carry any information about particular regions of space–time, but rather imposes certain global constraints on the space–time structure which have to be satisfied by every cosmological model purporting to describe the real Universe.

Here an interesting problem arises, namely the problem of the observational space–time. A space–time $(\mathcal{M}, g)$, being a solution of Einstein's field equations, is said to be *observational space–time associated with* $\mathcal{D}(w_0, z^*)$ if it contains a hypersurface Σ such that $\Sigma \cup \{q\}$ is the past light cone $L^-(q)$ of an event q situated on a regular timelike curve C, and if all observational data as predicted by the model $(\mathcal{M}, q)$

[10]This distance is determined by observing the pencil of null geodesics emanating from an observation event q. If this pencil forms a solid angle $d\Omega$ at q and spans a space element dS at a certain point p along the pencil, then the *area distance* from q to p is $dS = r^2 d\Omega$.

are exactly those data which are elements of $\mathcal{D}(w_0, z^*)$. If there exist more than one observational space–time associated with a given $\mathcal{D}(w_0, z^*)$, one is met with an *observational indeterminacy* of the model: there are more than one cosmological models that are consistent with observational data contained in $\mathcal{D}(w_0, z^*)$.

Actual achievements of the Ellis program consist in demonstrating that the maximal data set $\mathcal{D}(w_0, z^*)$ is simultaneously the smallest set of observational data required to uniquely determine the geometry of the past light cone $L^-(q)$ backward to $z = z^*$. In other words, the knowledge of the ideal data $\mathcal{D}(w_0, z^*)$ is the necessary and sufficient condition for uniquely determining, with the help of Einstein's equations, the geometry on our past light cone (backward to z^*). Doubtlessly, it is an interesting result. One cannot see any reason why it should be *a priori* true. One could easily imagine the situation in which a model of the Universe would be either undetermined or overdetermined by the set of ideal observational data available in principle (the first possibility occurs where the field equations are not taken into account; see above).

Geometric structure of the past light cone $L^-(\mathrm{q})$ is only a part of a cosmological model. Now, one should extend the solution obtained on the light cone to the Cauchy domain of dependence inside the cone. It is a difficult mathematical problem which can be successfully solved if relationships between mathematical quantities are analytic. However, there are some reasons suggesting that this assumption is not necessary. Since the analycity assumption is very restrictive, it would be desirable to have a proof without it.

It is also possible to treat $\mathcal{D}(w_0, z^*)$ as the initial data and look for a solution to the future, rather than backwards. This is a still more difficult problem. From the hyperbolic character of Einstein's equations it follows that one cannot expect a "deterministic" solution starting from a set of initial data, unless one assumes that the solution is analytic or that it cannot be perturbed by anything which could not be contained in the initial conditions (see sec. 1). The fact that, for a cosmologist, to reconstruct the past is more important than to forecast the future is hardly a consolation.

The question of stability of observational space–times is an important problem which still awaits its full elaboration. Roughly speaking, an observational space–time $(\mathcal{M}_0, g_0)$ associated with the maximal data set $\mathcal{D}_0(w_0, z^*)$ is said to be *observationally stable* if a small perturbation $\mathcal{D}(w_0, z^*)$ of the set $\mathcal{D}_0(w_0, z^*)$ leads to observational space–times $(\mathcal{M}_i, g_i)$, associated with $\mathcal{D}(w_0, z^*)$ such that $(\mathcal{M}_i, g_i)$ only slightly differ from $(\mathcal{M}_0, g_0)$. To define precisely the meaning of "slightly differs from" would require introducing a topology on the space of solutions to Einstein's equations. The observational stability condition should be satisfied by realistic world models. The situation in which slightly different observational data would lead to drastically different world models would be disastrous for any research in cosmology.

Although the analysis of the observational situation in cosmology with the ideal observations replaced by the real ones is still to be carried out, it is not difficult to foresee its general result. If the precise knowledge of the ideal data $\mathcal{D}(w_0, z^*)$ is neces-

sary and sufficient to reconstruct the geometry of our past light cone, and since there is no doubt that the postulate to be realistic will restrict the data (to the set $R(w_0, z^*)$, say), one can expect a significant observational indeterminacy of cosmological models. Analogously to the concept of observational space–times associated with $\mathcal{D}(w_0, z^*)$, one can introduce the concept of *realistic space–times associated with* $R(w_0, z^*)$. On the strength of the above analysis it is clear that the set of all realistic space–times associated with $R(w_0, z^*)$ consists of more than one space–time. This statement can be thought of as a formulation of the *observational indeterminacy principle* in cosmology. It turns out to be an unavoidable consequence of the extrapolation method used in cosmological research.

Observational indeterminacy in cosmology can be either acknowledged as a necessary element of our world picture, or neutralized with the help of different philosophical postulates such as the cosmological principle claiming that the earthly observer occupies a typical position in the Universe (see Ellis 1975; Heller 1978). If one chooses the second possibility, one has to define cosmology not only as a science about the Universe, but also as a science about the assumptions which must be made in order to render such a science possible. If we take into account the fact that even the observational Ellis' program is based on many (both explicit and tacit) assumptions (laws of physics are everywhere the same, space–time is modelled by the pair $(\mathcal{M}, g)$, etc.), it seems that the above definition of cosmology can hardly be avoided.

Chapter 6

Geometry of the Standard Cosmological Model

0. Introduction

In the preceding chapter we have discussed some fundamental problems of cosmology understood as a non local physics. The discussion gave us a kind of program or suggested a style of research in cosmology. Unfortunately, the scope of the present book does not allow us to implement this challenge in a more ambitious way. Especially, to deal with cosmology as a theory of the ensemble of universes would require a separate monograph. Moreover, it would seem more reasonable to postpone writing such a monograph till the domain is worked out more fully. Nevertheless, the present book cannot be complete without at least touching the space of solutions of Einstein's equations. No wonder that in such a situation our choice goes to a subset (of the zero measure!) of this space, namely to the *Robertson–Walker solutions*. These solutions are distinguished: first, by the maximal degree of symmetry admitting the existence of the "open" global time; second, by the surprising property that there are exactly these simplest, in a sense, world models that correctly — albeit in an approximate way — describe the observed Universe. Because of the last property one sometimes speaks of the *standard cosmological model*.

The present chapter is intended as an example of doing cosmology according to the program outlined in the preceding chapter (inasmuch as a zero measure set can serve as an example of anything!). Therefore, from the very beginning we turn to global mathematical methods. They allow us to compactify the material and pay off with precise definitions of such concepts as particle with a non–zero rest–mass, photon, observer, etc. which have to be put into the formalism in order to change mathematics into cosmology. The books by Sachs and Wu (1977) and by O'Neill (1983) were both good examples and a great help. However, in the final parts of the chapter, concerning observational aspects of cosmology, we postpone this formalistic approach. It turns out to be too narrow for embracing the observational practice. To treat a galaxy or a cluster of galaxies as a curve in space time can be very precise from the mathematical point of view, but — if taken literally — does not allow one even meaningfully to speak about the angular diameter of a celestial object. Cosmology, in a very transparent way, reveals both the weakness and the strength of the empirical method: The weakness — since the reality, as it manifests itself in empirical results, turns out to be too rich to be translated into the available

mathematical structures; and the strength — since simple mathematical structures surprisingly well approximate the reality.

The organisation of the present chapter is the following. In sec. 1, the concept of warped (Cartesian) product is introduced; it turns out to be a good tool significantly simplifying subsequent analyses. In sec. 2, the aforementioned definitions of particles and observers are presented, and some propositions connected with them are proved; they will serve us afterwards in shortening computations. The construction of the standard cosmological model (of the Robertson–Walker space–time) is carried out in sec. 3. The most important features of the evolution of the standard world model (singularities, spatial curvature, etc.) are expressed in the three theorems. The goal of sec. 4 is to formulate and to prove them. In cosmological formulations two subclasses of the Robertson–Walker world models play an important role, namely dust– and radiation–filled universes. Solutions describing these models are presented in sec. 5. Observational aspects of the standard world model are (very) briefly discussed in sec. 6. Many comments and remarks can be made as far as the standard model is concerned; in sec. 7, we focus on the problem of idealisations that are incorporated into the construction of this model; we briefly mention solutions with the cosmological constant different from zero and world models with non–standard topologies (which are locally Robertson–Walker space–times); we touch upon a difficult problem concerning causally disconnected regions of space–time and the two attempts to solve it: the older one, known under the name of the mixmaster mechanism, and the newer one, the so–called inflationary scenario.

The present chapter is followed by a *Bibliographical Essay*; its aim is to introduce the reader, who might feel disappointed with the brevity of the chapter dealing with the standard universe, to a fascinating realm of books on cosmology and related topics.

1. Warped Products

Let us consider a Cartesian product $\mathcal{B} \times \mathcal{F}$ of two pseudo–Riemannian manifolds $(\mathcal{B}, g_B)$ and $(\mathcal{F}, g_F)$, where g_B and g_F are metric tensors on $\mathcal{B}$ and $\mathcal{F}$, correspondingly. $\pi^*(g_B) + \sigma^*(g_F)$, where $\pi: \mathcal{B} \times \mathcal{F} \to \mathcal{B}$, $\sigma: \mathcal{B} \times \mathcal{F} \to \mathcal{F}$ are projections, is a metric tensor on $\mathcal{B} \times \mathcal{F}$. Let $f > 0$ be a smooth function on $\mathcal{B}$.

1.1. Definition. Cartesian product $\mathcal{B} \times \mathcal{F}$, equipped with the metric tensor

$$g = \pi^*(g_B) + (f \circ \pi)^2 \sigma^*(g_F),$$

is called *warped product* and denoted by $\mathcal{M} = \mathcal{B} \times_f \mathcal{F}$.

If $f = 1$, the warped product is the usual Cartesian product. As it can be easily seen, warped product is a special case of a fibre bundle. Therefore, it is reasonable to call $\mathcal{B}$ a *base manifold* (or *base*, for short), and $\pi^{-1}(p) = p \times \mathcal{F}$, $p \in \mathcal{B}$, a fibre over p

(fig. 6.1). Fibres and cross-sections $\sigma^{-1}(q) = \mathcal{B} \times q$, $q \in \mathcal{F}$, are pseudo-Riemannian manifolds. f is said to be a *warping function.*

1.2. Example. All surfaces of revolution are warped products. Let $\mathcal{M}$ be a surface obtained by the revolution of a flat curve C around an axis in $\mathbf{R}^3$. The distance from the curve C to this axis is given by the function $f: C \to \mathbf{R}^1$. One has $\mathcal{M} = C \times_f \mathcal{S}^1$.

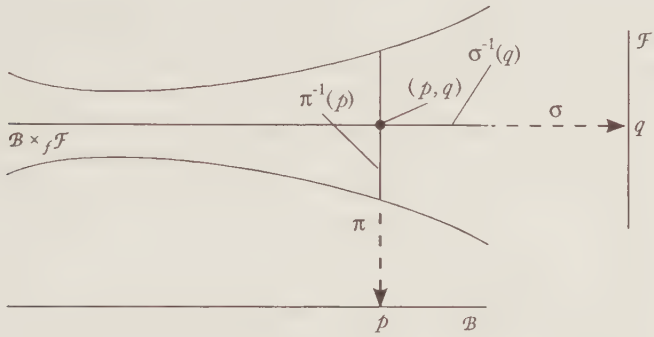


Fig. 6.1. Warped product $\mathcal{M} = \mathcal{B} \times_f \mathcal{F}$.

1.3. Definitions. Let $\mathcal{B} \times \mathcal{F}$ be a Cartesian product of manifolds $\mathcal{B}$ and $\mathcal{F}$, π and σ the projections as above, and f a function on $\mathcal{B}$. The function $\tilde{f} = f \circ \pi$ defined on $\mathcal{B} \times \mathcal{F}$ is said to be the *lift of the function* f to the product $\mathcal{B} \times \mathcal{F}$. Let $p \in \mathcal{B}$, $q \in \mathcal{F}$, and $v \in \mathcal{T}_p(\mathcal{B})$. The unique vector $\tilde{v} \in \mathcal{T}_{(p,q)}(\mathcal{B} \times \mathcal{F})$, such that $d\pi(\tilde{v}) = v$, is said to be the *lift of the vector* v to $(p, q) \in \mathcal{B} \times \mathcal{F}$.

In an analogous manner, one can *lift* functions and vectors from the manifold $\mathcal{F}$ to $\mathcal{B} \times \mathcal{F}$. The above definitions remain valid also with respect to warped products.

1.4. Definition. Let $(\mathcal{M}, g_M)$ and $(\mathcal{N}, g_N)$ be pseudo Riemann manifolds. A diffeomorphism $\psi: \mathcal{M} \to \mathcal{N}$ such that $\psi^*(g_N) = cg_M$, where $c = \text{const} \neq 0$, is called a *homotetic mapping*, or a *homotety* for brevity (it is a special case of a *conformal mapping* which is defined by the condition: $c = h$ is a different from zero function on $\mathcal{M}$). $|c|^{\frac{1}{2}}$ is called a *scale factor.*

It is easy to prove the following.

1.5. Proposition. Let $\mathcal{M} = \mathcal{B} \times_f \mathcal{F}$ be a warped product,

(a) for every $q \in \mathcal{F}$, the mapping $\pi|_{(\mathcal{B} \times q)}$ is an isometry on $\mathcal{B}$;
(b) for every $p \in \mathcal{B}$, the mapping $\sigma|_{(p \times \mathcal{F})}$ is a homotety with the scale factor $1/f(p)$;
(c) for every $(p, q) \in \mathcal{M}$, the fibre $\pi^{-1}(p)$ and the cross-section $\sigma^{-1}(q)$ are orthogonal at (p, q). □

The properties expressed in this proposition allow one easily to reconstruct the geometry of the warped product $\mathcal{B} \times_f \mathcal{F}$ when the geometry is known of

the manifolds $\mathcal{B}$ and $\mathcal{F}$ (see O'Neill 1983, pp. 204–211; Beem and Ehrlich 1981, pp. 55–79).

The concept of warped product was introduced by Bishop and O'Neill(1969).

2. Macroscopic Description of Particles and Observers

Let us begin with definitions the aim of which is to make the description more precise of an "averaged matter content" of the Universe traditionally used in cosmology. In what follows, all structures are defined on the space–time manifold $(\mathcal{M}, g)$.

2.1. Definition. A timelike future–oriented curve $\alpha: I \to \mathcal{M}$, such that $|\alpha'(\tau)| = 1$ for all $\tau \in I$, is said to be an *observer* in space–time $\mathcal{M}$; τ is his proper time.

2.2. Definition. A timelike future–directed curve $\alpha: I \to \mathcal{M}$, such that $g(\alpha', \alpha') \equiv \langle \alpha', \alpha' \rangle = -m^2$, is said to be a *particle with the rest–mass* m, or a *particle* for brevity.

2.3. Definition. If the curve α of definition 2.1 or 2.2 is a geodesic curve, the observer or the particle is called *freely falling*.

2.4. Definition. A future–directed zero geodesic $\gamma: I \to \mathcal{M}$ is said to be a *photon* or a *light beam*.

Images $\mathrm{Im}(\alpha)$, $\mathrm{Im}(\gamma) \subset \mathcal{M}$ of the curves $\alpha: I \to \mathcal{M}$ and $\gamma: I \to \mathcal{M}$ are one–dimensional manifolds; they are called *histories* (or *world lines*) of observers or particles.

2.5. Definitions. Timelike future–directed unit vector $v \in \mathcal{T}_q(\mathcal{M})$ is said to be an *instantaneous observer* at $q \in \mathcal{M}$. By performing the orthogonal decomposition of the tangent space $\mathcal{T}_q(\mathcal{M}) = \mathbf{R}u + u^\perp$, one obtains the *time axis* $\mathbf{R}u$ of the observer and its *rest–space* $u^\perp$.

One could imagine that an instantaneous observer is situated in $\mathcal{T}_q(\mathcal{M})$ and observes the Universe with the help of the *exponential mapping* $\exp_q: \mathcal{T}_q(\mathcal{M}) \to \mathcal{M}$. If $u \in \mathcal{T}_q(\mathcal{M})$, then there exists a unique geodesic curve α such that $a'(0) = u$; the exponential mapping is defined by $\exp_q(u) = \alpha(1)$. By the phrase "to observe the space–time $\mathcal{M}$ from the tangent space $\mathcal{T}_q(\mathcal{M})$ with the help of the exponential mapping" one should understand projecting structures from an open neighbourhood $U \subset \mathcal{M}$ of q by $\exp_q^{-1}$ and treating them as structures on $\mathcal{T}_q(\mathcal{M})$ (see Sachs and Wu 1977a, sec. 4.2).

2.6. Definition. A unit vector field, the integral curves of which are observers (of definition 2.1), is said to be a *reference frame* in the space–time $(\mathcal{M}, g)$. Observers in question will be called *observers in a given reference frame*.

Satisfactory agreement with actual observational data is achieved under a surprisingly strong simplification allowing one to treat the "material content of the Universe" as a perfect fluid. In physics, perfect fluid is defined as a continuous medium moving, at each point $q \in \mathcal{M}$, with a velocity v_q such that an observer (in an every–day meaning of this term), moving with the same velocity v_q, sees this medium as

isotropically distributed in space. Till the end of the present section, all analyses will be carried out locally, i.e., in a neighbourhood of a point q at which the perfect fluid approximation remains valid. The above ideas are formalized in the following.

2.7. Definition. A *perfect fluid* in space-time $(\mathcal{M}, g)$ is the triple (U, ρ, p), where U is a reference frame, called the *co-moving reference frame*; $\rho\colon \mathcal{M} \to [0, \infty)$ is the so-called *energy density function*, and $p\colon \mathcal{M} \to [0, \infty)$ is the so-called *pressure* (or *stress*) *function* (it is assumed that ρ and p are of C^∞ class). *The energy-momentum tensor of the perfect fluid* is defined to be

$$T = (\rho + p)U^* \otimes U^* + pg, \tag{2.1}$$

where U^* is a one-form metrically equivalent to the vector field U.

2.8. Proposition. The components of the energy-momentum tensor of the perfect fluid (in the co-moving reference frame) are

$$T_{00} = \rho, \qquad T_{ij} = p\delta_{ij}, \quad (i, j = 1, 2, 3). \ \blacksquare$$

Proof. Let U be the co-moving frame. We shall use the basis (U, X, Y) where $X, Y \perp U$. Then $T(U, U) = \rho$; $T(U, X) = 0$; $T(X, X) = \langle X, Y \rangle$. □

If $p = 0$, the perfect fluid is called *dust*; if $p = \frac{1}{3}\rho$, the perfect fluid is called *photon gas* or *radiation field*.

The local conservation laws requires $\operatorname{div} T = 0$.

2.9. Proposition. Let (U, ρ, p) be a perfect fluid. $\operatorname{Div} T = 0$ implies

(1) $$U\rho = -(\rho + p)\operatorname{div} U;$$

(2) $$-(\operatorname{grad} p)^{\perp} = (\rho + p)\nabla_U U. \ \blacksquare$$

Proof. Let ${}^*T$ be a tensor field of the type $(2, 0)$ metrically equivalent to the tensor field T. One has $(\operatorname{div} T = 0) \Rightarrow (\operatorname{div} {}^*T = 0)$, and one obtains

$$\operatorname{div} {}^*T = U(\rho + p)U + (\rho + p)\nabla_U U + (\rho + p)(\operatorname{div} U)U + \operatorname{grad} p = 0.$$

The vertical component of this equation $(\langle \operatorname{div} {}^*T, U \rangle = 0)$ gives equation (1), and its longitudinal component $(\langle \operatorname{div} {}^*T, \nabla_U U \rangle = 0)$ equation (2). □

Equation (1) expresses the rate of change of the energy density as measured by an observer in the co-moving reference frame U. If in equation (2) the space component of the gradient of the pressure function $((-\operatorname{grad} p)^{\perp})$ is treated as a force, the coefficient $(\rho + p)$ as a counterpart of mass, and $\nabla_U U$ as an acceleration, then equation (2) turns out to be Newton's second law.

The presented here formalised macroscopic description of particles and observers was originally proposed by Sachs and Wu (1977a, b), and its somewhat simplified form can be found in O'Neill (1983) (see also Heller 1982).

3. Construction of the Standard Cosmological Model

In a geometric approach to physics adopted also in the present book — in contradistinction to the purely mathematical texts — definitions are more important than theorems. Very often definitions express fundamental facts or experimental generalizations, whereas theorems are usually only their formal consequences. Through definitions the reality enters mathematics (see Sachs and Wu 1977a, sec. 1.5). The following definition constitutes both a mathematical stylization and a synthesis of many observational facts lying at the basis of the standard cosmological model (these facts will be briefly presented in sec. 6 of the present chapter).

3.1. Definition. Let $\mathcal{S}$ be a connected three–dimensional Riemann manifold with constant curvature $k = -1, 0$, or $+1$; I be an open interval of the space $\mathbf{R}^1_1$, where $\mathbf{R}^1_1$ denotes the space $\mathbf{R}^1$ with the signature 1; $R(t) > 0$, $t \in I$ be a smooth real function on I. The warped product $\mathcal{M} = I \times_R \mathcal{S}$ is called a *Robertson–Walker space–time*.

A Robertson–Walker space–time is, therefore, a manifold $I \times \mathcal{S}$ with the line element

$$ds^2 = -dt^2 + R^2(t)d\sigma^2 \tag{3.1}$$

where $d\sigma^2$ is a line element on the manifold $\mathcal{S}$ lifted to $I \times \mathcal{S}$. Fibres $\mathcal{S}(t)$, $t \in I$, are called *instantaneous spaces* or *space sections* (at an instant t) of the Robertson–Walker space–time. Every instantaneous space $\mathcal{S}(t)$ has a constant curvature equal to $k/R^2(t)$. The warping function $R(t)$, in this context, will be called the *scaling function*, and its value $R(t_0)$, for an instant $t_0 \in I$, the *scale factor*.

The line element σ^2 can be reduced to the one of the following two forms,

$$d\sigma^2 = \frac{(dx^1)^2 + (dx^2)^2 + (dx^3)^2}{\left\{1 + \frac{k}{4}\left[(x^1)^1 + (x^2)^2 + (x^3)^2\right]\right\}^2} \tag{3.2a}$$

$$= \frac{dr^2 + r^2 d\theta^2 + r^2 \sin^2\theta d\varphi^2}{\left[1 + \frac{kr^2}{4}\right]^2} \tag{3.2b}$$

Let d/dt be a vector field on $I \subset \mathbf{R}^1$, and $U = \partial/\partial t$ its lifting to $I \times \mathcal{S}$. The curve $I \times q$, $q \in \mathcal{S}$, can be parametrized in the following way: $\alpha_q(t) = (t, q)$. Of course, it is an integral curve of the field U, and consequently an observer in the reference frame U (definition 2.6); such an observer will be called a *fundamental observer*. Thus, the function t ascribes to every fundamental observer his proper time; this function will be called the *cosmic time function*, or *time function*, for brevity.

It can be easily checked that $\langle U, U\rangle = g(U, U) = -1$, and $U \perp \mathcal{S}(t)$, for every $t \in I$ (see proposition 1.5c), i.e., every instantaneous space $\mathcal{S}(t)$ is a spacelike hypersurface in the space–time $\mathcal{M} = I \times_R \mathcal{S}$.

Let us introduce two further concepts playing important roles in cosmology.

3.2. Definition. A space–time $\mathcal{M} = I \times_R \mathcal{S}$ is said to be *spatially isotropic with respect to a point* $q \in \mathcal{S}(t)$ if the point $(t, q) \in \mathcal{M}$ has a neighbourhood $\mathcal{N}$ such that,

for any two unit vectors v and w tangent to $\mathcal{S}(t)$ at (t, q), there exists an isometry $\phi: \mathcal{N} \to \mathcal{N}$ such that $\phi = \mathrm{id} \times \phi|_{\mathcal{S}}$ and $d\phi(v) = w$.

3.3. Definition. A space–time $\mathcal{M} = I \times_R \mathcal{S}$ is said to be *spatially homogeneous* if, for $p, q \in \mathcal{S}(t)$, $t \in I$, there exists an isometry $\psi: \mathcal{U} \to \Theta$, where $\mathcal{U}$ and Θ are neighbourhoods of p and q respectively.

Less strictly but quite commonly, one simply says that the space $\mathcal{S}(t)$ is isotropic around a point $q \in \mathcal{S}(t)$ or homogeneous.

The concept of spatial isotropy formalises the intuitive statement that there are no privileged directions with respect to a point q. The concept of spatial homogeneity formalizes the intuitive statement that there are no privileged points in the considered space. These concepts can be more generally defined in terms of the Lie group theory (see Wolf 1967, Robertson and Noonan 1968). It is known from the Schur theorem that a space which is isotropic with respect to each of its points is also homogeneous and that a space is isotropic with respect to each of its points if and only if it is a space of constant curvature (see Wolf 1967, p. 57).

From these considerations we obtain the following.

3.4. Corollary. Instantaneous spaces of a Robertson–Walker space–time are isotropic around every of its points, and consequently they are also homogeneous. □

This corollary plays an important role in observational motivation of the the choice of a Robertson–Walker space–time for the standard cosmological model.

3.5. Proposition. For any Robertson–Walker space–time $\mathcal{M} = I \times_R \mathcal{S}$, in the reference frame $U = \partial/\partial t$, non-vanishing components of the Ricci tensor are

$$\mathfrak{R}_{00} = -3\frac{\ddot{R}}{R},$$

$$\mathfrak{R}_{11} = \mathfrak{R}_{22} = \mathfrak{R}_{33} = -2\frac{\dot{R}^2}{R^2} + 2\frac{k}{R^2} + \frac{\ddot{R}}{R}$$

(notice the change of the Ricci tensor symbol as compared with chapter 4) and the curvature scalar is

$$\mathfrak{R} = 6\left\{\frac{\dot{R}^2}{R^2} + \frac{k}{R^2} + \frac{\ddot{R}}{R}\right\}. \blacksquare$$

Proof by direct computation from the definition; one can also employ the theorem on the Riemann and Ricci tensors in warped products (see O'Neill 1983, pp. 209–211). □

Einstein's equations introduce dynamics into the cosmological model. In the following, for the sake of simplicity, we will use Einstein's equations in the form (5.7) of chapter 4, i.e., without the cosmological constant. This simplification is motivated by observational data which point out that cosmological constant effects at the present epoch are negligible within limits of observational accuracy (see Petrosian 1974). In the following, we shall use the *geometric system of units* in which the light velocity c

and the Newtonian constant of gravity G are assumed to be equal to one; in such a case the constant κ appearing in equation (5.7) of chapter 4 has the value 8π.

3.6. Theorem. Perfect fluid (U, ρ, p) in the Robertson–Walker space–time $\mathcal{M} = I \times_R \mathcal{S}$ is described by the following equations

$$\frac{8\pi}{3}\rho = \frac{\dot{R}^2}{R^2} + \frac{k}{R^2}, \tag{3.3}$$

$$-8\pi p = 2\frac{\ddot{R}}{R} + \frac{\dot{R}^2}{R^2} + \frac{k}{R^2}. \blacksquare \tag{3.4}$$

Proof by direct computation from Einstein's equations $\mathfrak{R}_{\mu\nu} - \frac{1}{2}\mathfrak{R}g_{\mu\nu} = -8\pi\mathcal{T}_{\mu\nu}$; proposition 3.5 can be used. By comparing equations (3.3) and (3.4) with formula (2.1) one can see that these equations describe a perfect fluid. □

Definition 3.1, together with theorem 3.6 constitutes the essence of the construction of the standard cosmological model.

3.7. Corollary.

$$-4\pi(\rho + 3p) = -3\frac{\ddot{R}}{R}. \square \tag{3.5}$$

3.8. Corollary. Equation (1) of proposition 2.9 in a Robertson–Walker space–time assumes the form

$$U\rho = \dot{\rho} = -3(\rho + p)\frac{\dot{R}}{R}. \tag{3.6}$$

(Equation (2) of the same proposition is trivially satisfied since in the considered case $(\operatorname{grad} p)^{\perp} = \nabla_U U = 0$.) □

4. Evolution of the Standard Cosmological Model.

The structure of the cosmological model presented in the preceding section is rich enough to contain in itself information concerning the world evolution and its limitations. This information is deciphered in the following theorems.

4.1. Theorem. (O'Neill 1983, p. 348). Let $\mathcal{M} = I \times_R \mathcal{S}$ be a Robertson–Walker space–time satisfying the condition: $\rho + 3p > 0$ (the *strong energy condition*). If for a certain $t_0 \in I$ the relationship $H_0 \equiv H(t_0) \equiv \dot{R}_0/R_0 > 0$ is satisfied, then the interval I has the initial point t_{in} such that $t_0 - H_0^{-1} < t_{\text{in}} < t_0$, and either (1) $\dot{R} > 0$ or (2) R attains maximum at $t_{\max}$, $t_{\max} > t_0$. If (2) takes place, then $I = (t_{\text{in}}, t_{\text{fin}})$. ■

Proof. From the condition $\rho + 3p > 0$ and relationship (3.5) it follows that $\ddot{R} < 0$, i.e., that the diagram of the function $R(t)$ is a convex curve. In particular, this diagram is everywhere (besides the tangency point) beneath the diagram of the function

$$\mathcal{Q}(t) = R(t_0) + H_0 R(t_0)(t - t_0),$$

which is the tangent line to $R(t)$ at the point $R(t_0)$. The condition $H_0 > 0$ implies vanishing of the function $Q(t)$ at $t_0 - H_0^{-1}$ (see, fig. 6.2). Since, however, the diagram of the function $R(t)$ is a convex curve and $R(t)$ cannot assume negative values, $R(t)$ must break down at a certain instant t_{in} situated between $t_0 - H_0^{-1}$ and t_0 .

From the fact that $\ddot{R}(t) < 0$ on the entire I, it follows that either (1) $\dot{R} > 0$ on the entire I or (2) R attains maximum for a certain $t_{\max} > t_0$, and afterwards $\dot{R} < 0$. In the last case, the reasoning analogous to that in the first part of the proof leads to the conclusion that I must have the final point $t_{\text{fin}} > t_0$. □

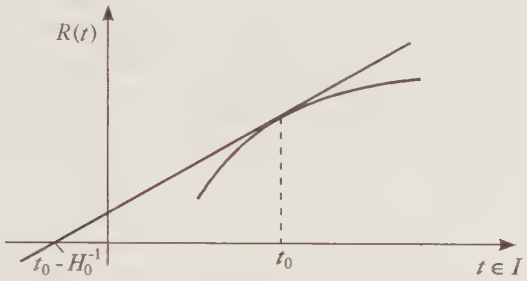


Fig. 6.2. Breaking down of the history of the Robertson–Walker world model at the instant t_{in}; $t_0 - H_0^{-1} < t_{\text{in}} < t_0$

Note that theorem 4.1 says nothing about the state of the Universe in the process of its approaching backward in time to the initial instant t_{in}; in particular, it does not say that $\rho \to \infty$ and $R \to 0$ when $t \to t_{\text{in}}$.

4.2. Theorem. Let $\mathcal{M} \times_R \mathcal{S}$ be a Robertson–Walker space-time with $H_0 > 0$ for a certain $t_0 \in I$. If $\rho > 0$, $-\frac{1}{3} < a \leq p/\rho$, $a = \text{const}$, and $R \to 0$ as $t \to t_{\text{in}}$, then $\dot{R} \to \infty$ and $\rho \to \infty$ when $t \to t_{\text{in}}$. ■

Proof. From the assumption $-\frac{1}{3} < a \leq p/\rho$ it follows that $\rho + 3p > \varepsilon p > 0$ for a certain $\varepsilon > 0$. Therefore, theorem 4.1 applies; in particular $\ddot{R} < 0$.

Since in the interval (t_{in}, t_0) one has $\dot{R} > 0$, and on the strength of the above condition $\rho - \varepsilon\rho \geq -3p$, equation (3.4) changes into the inequality

$$\dot{\rho} \leq -(2+\varepsilon)\rho\frac{\dot{R}}{R},$$

which, as can be easily seen, is equivalent to

$$\frac{d}{dt}(\rho R^{2+\varepsilon}) \leq 0,$$

i.e., in the interval (t_{in}, t_0) one has $\rho R^{2+\varepsilon} \geq \rho(t_0)R(t_0)^{2+\varepsilon}$. However, from the assumption one knows that $R \to 0$ if $t \to t_{\text{in}}$. Therefore, $\rho R^2 \to \infty$ and $\rho \to \infty$ as $t \to t_{\text{in}}$.

Equation (3.3) can be given the form

$$\frac{8\pi}{3}\rho R^2 = \dot{R}^2 + k, \tag{4.1}$$

which immediately leads to the conclusion that also $\dot{R} \to \infty$ as $t \to t_{\text{in}}$. □

Therefore, the evolution of the standard cosmological model begins with the singularity, the so–called *initial singularity*. It is characterized by the fact that as time, going backwards, approaches the initial instant, $t \to t_{in}$, both the energy density ρ and the expansion rate $\dot{R}$ grow indefinitely. This fact justifies the name *Big Bang*.

4.3. Theorem. Let $\mathcal{M} = I \times_R \mathcal{S}$ be a Robertson–Walker space–time satisfying the conditions of theorems 4.1 and 4.2. (A) If $\dot{R} > 0$ on the entire interval $I = (t_{\text{in}}, \infty)$, then either $k = 0$ or $k = -1$. (B) If R attains maximum at $t_{\max} \in I$, then $k = +1$ (and $I = (t_{\text{in}}, t_{\text{fin}})$), and if $R \to 0$, as $t \to t_{\text{fin}}$, then $\rho \to \infty$ and $\dot{R} \to -\infty$, as $t \to t_{\text{fin}}$.■

Proof. (A) First, let us assume that $\dot{R} > 0$ on the entire interval $I = (t_{\text{in}}, \infty)$. As in the preceding theorem, $p > 0$ and $\rho + 3p > 0$, which implies $\rho + p > 0$. Therefore, equation (3.6) leads to the conclusion that $\dot{\rho} < 0$, i.e., ρ decreases indefinitely. Two cases ought to be considered:

(1) $R \to \infty$ as $t \to \infty$. From the proof of theorem 4.2 it is known that $\frac{d}{dt}(\rho R^{2+\varepsilon}) \leq 0$ (but now on the entire I) which implies that $\rho R^{2+\varepsilon}$ is a bounded function for large $t \in I$. In other words, $\rho R^2 \to 0$ as $t \to \infty$. From equation (4.1) one gets either $k = 0$ or $k = -1$.

(2) However, the case $R \to b = \text{const}$ as $t \to \infty$ cannot be excluded. With this assumption, and remembering that $\ddot{R} < 0$, one obtains that $\dot{R} \to 0$. This, together with equation (4.1) leads to the conclusion that either $k = 0$ or $k = +1$. However, the case $k = +1$ is excluded. Indeed, let us assume, on the contrary, that $k = +1$. From equation (4.1) one gets $\rho R^2 \to 3/(8\pi)$, and consequently $\rho \geq \delta$ for a certain $\delta > 0$. Since $\dot{R} \to 0$ as $t \to \infty$, there exists a sequence $\{t_i\} \to \infty$, $t_i \in I$, such that $\{\ddot{R}(t_i)\} \to 0$. Therefore, equation (3.5) gives $\{(\rho + 3p)(t_i)\} \to 0$, but this contradicts the inequality $\rho + 3p > \varepsilon p > 0$ (see, beginning of the proof of theorem 4.2).

Therefore, if $\dot{R} > 0$ on the entire I, then either $k = 0$ or $k = -1$. (It can be shown that the case $R \to b$ as $t \to \infty$ also excludes $k = 0$; see O'Neill 1983, p. 349.)

(B) Let us assume that R attains maximum at $t_{\max} \in I$, i.e., $\dot{R}(t_{\max}) = 0$. Equation (3.3) gives $\rho(t_{\max}) = 3k/8\pi R^2(t_{\max}) > 0$; hence $k = +1$. The rest of the proof is similar to that of theorem 4.2. □

4.4. Corollary. Let us define the so–called *critical density* $\rho_{\text{crit}} = 3H_0^2/8\pi$. If $\rho_0 = \rho_{\text{crit}}$, then $k = +1$ (index zero denotes the value of the corresponding parameter at a certain instant $t_0 \in I$, for example at the instant of performing observation). ■

Proof. From equation (3.3) one gets $\rho_0 - \rho_{\text{crit}} = 3k/8\pi R_0^2$. □

5. Dust Filled and Radiation Filled Universes

Two special cases of perfect fluid Robertson–Walker space–times play an important role in cosmological applications, namely dust filled and radiation filled world models. We will consider them in turn.

5.1. Proposition. The scaling function $R(t)$ of a Robertson–Walker space–time $\mathcal{M} = I \times_R \mathcal{S}$ filled with dust $(U, \rho, p = 0)$ satisfies the so–called *Friedman equation*

$$\dot{R}^2 = \frac{8\pi}{3}\frac{E}{R} - k, \tag{5.1}$$

where $E = \rho R^3 > 0$ is a conserved magnitude on the entire interval I. ■

Proof. Equation (5.1) is equivalent to equation (3.3). Equation (3.6) with $p = 0$, after simple transformations, gives $d(\rho R^3)/dt = 0$. □

Solutions of the Friedman equation describe the evolution of the cosmological model. In the considered case, these solutions — for different values of the parameter k — are the following,

$$\begin{aligned}
&k = 0: && R = Ct^{2/3}, \qquad C = 6\pi E;\\
&k = +1: && t = \tfrac{4}{3}\pi E(\theta - \sin\theta);\\
& && R = \tfrac{4}{3}\pi E(\theta - \cos\theta), \qquad (0 < \theta < 2\pi);\\
&k = -1: && t = \tfrac{4}{3}\pi E(\sinh\eta - \eta);\\
& && R = \tfrac{4}{3}\pi E(\cosh\eta - 1) \qquad (\eta > 0).
\end{aligned}$$

These solutions are shown in fig. 6.3; they are often called *dust filled Friedman universes, or Friedman universes* for short. Their evolution starts with the Big Bang (the instant t_{in} of the Big Bang has been accepted as the beginning of time counting, i.e., $t_{\text{in}} = 0$). In the case $k = 0$ (the *flat model*), $R \to \infty$ and $\dot{R} \to 0$ as $t \to \infty$. This solution is also called the *Einstein–de Sitter model*. If $k = +1$, the solution is a branch of the cycloid. The maximum occurs at $t_{\max} = (4/3)\pi^2 E$ and is $R_{\max} = 8\pi E/3$. The final singularity takes place at $t_{\text{fin}} = 8\pi^2 E/3$. In the case $k = -1$, $R \to \infty$ and $\dot{R} \to 1$ as $t \to \infty$.

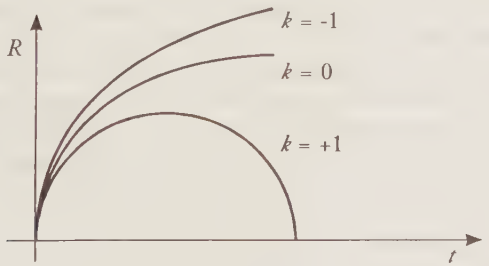


Fig. 6.3. Friedman's world models.

5.2. Proposition. The scaling function R(t) of a Robertson–Walker space–time $\mathcal{M} = I \times_R \mathcal{S}$ filled with radiation $(U, \rho, p = \rho/3)$ satisfies the equation

$$\dot{R}^2 = \frac{8\pi\varepsilon}{3R^2} - k, \tag{5.2}$$

where $\varepsilon = \rho R^4$ is a conserved magnitude on the entire interval I. ■

Proof. Equations (3.4) and (3.5), after straightforward manipulations taking into account that $p = \rho/3$, give equation (5.2). In this case, equation (3.6) turns out to be equivalent to $d(\rho R^4)dt = 0$. □

The solutions of equation (5.2) are the following,

$$\begin{aligned} k = 0: &\qquad R = (2bt)^{\frac{1}{2}}; \\ k = +1: &\qquad R = (2bt - t^2)^{\frac{1}{2}}; \\ k = -1: &\qquad R = (2bt + t^2)^{\frac{1}{2}}. \end{aligned}$$

where $b = (8\pi\varepsilon/3)^{\frac{1}{2}}$.

5.3. Corollary. If $t \to t_{\text{in}}$, all solutions behave as the flat solution, i.e., curvature effects become negligible. □

6. Observational Background of the Standard Cosmological Model

So far analyses of the present chapter had a purely mathematical character. Definitions of observers, particles, perfect fluid etc. (see, sec. 2) were very formal and rather loosely connected with physical or astronomical practice. Now, the formalism must be put into resonance with the results of measurements and observations. In other words, mathematical definitions, theorems and corollaries must be changed into what is usually called a cosmological model. These procedure immediately relaxes the entire formal fabric. Even a high idealisation of physical objects (e.g., of galaxies to points in space or to curves in space–time, see below) impresses on mathematics a stigma of empirical indeterminacy. Therefore, in the present section we postpone our previous mathematical style of exposition and change into a more "narrative language". Consequently, terms such as observer, particle, and so on, will be understood in their intuitive everyday meaning, and not in the sense of definitions 2.1–2.7.

First of all, one should be aware of the kind of vicious circle in which every empirical testing of cosmological models is entangled (in fact, it is a feature of the empirical method as such, but in cosmology it could be more dangerous than in other branches of the empirical sciences). On the one hand, without the help of a theory or of a model one does not know what and how to measure, and in which way the obtained "data" should be interpreted. On the other hand, with no observational data one could hardly create a theory or a model, not to mention testing of their validity. This vicious circle has to be somehow broken down. The method — honoured by

practice and supported by surprising, albeit not quite understandable, successes — proceeds in the following way. At the starting point one accepts (usually tacitly) the simplest model; the simplest in this context means remaining in agreement with a scientific tradition or with intellectual habits. Conceptual tools such a model supplies allow one to gather and to interpret observational data. As long as consistency of interpreted data with the model can be maintained, there is no reason to look for another model. Only under the pressure of increasing inconsistencies and anomalies, scientists begin (not without a certain unwillingness) to revise the model, to change some of its assumptions, to revaluate some of its argumentation methods ... This process usually leads to a new model which guarantees, at least for the time being, agreement of the theory with observational data. If necessary, the history repeats itself.

Cosmology of the twentieth century began its evolution from a traditional and intuitive model of a static flat and Euclidean space and an absolute time. Scientists believed that such a model was implied by the Newtonian theory of gravity (or at least it did not contradict the Newtonian theory) and remains in agreement with astronomical observations.[1] Gradually, observational data discarded this model. The standard cosmological model emerged as a better approximation to the reality; within its framework a satisfactory agreement has been reached between the theory and observation. However, it cannot be thought of as "empirically proved" but only as a well–functioning model providing a consistent interpretation of empirical data. Exclusively in this sense, in the following we shall speak about observational verification of the standard world model.

First of all, one must observationally verify definition 3.1 which constitutes the core of the standard model. If this definition is to be referred to the physical reality, the space–time $\mathcal{M}$ of the Universe is supposed to be of the form of the warped product $\mathcal{M} = I \times_R \mathcal{S}$, where $\mathcal{S}(t)$ are instantaneous spaces of constant curvature, and the parameter $t \in I$ represents the cosmic time. The empirical motivation for the isotropy and homogeneity of instantaneous spaces is provided by (A) measurements of the microwave background radiation and (B) investigations of the spatial distribution of galaxies. The information that $R(t) \neq \text{const}$ is supplied by (C) red shift measurements in galactic spectra. The family of three–dimensional homogeneous and isotropic spaces[2] can be, in a natural way, parametrized by the parameter $t \in I$; this defines the projection $\pi: \mathcal{M} \to I$ which can be interpreted as the existence of the cosmic time in the model (see Appendix). In what follows we shall notice some difficulties which this apparently simple model encounters.

(A) The microwave background radiation has the following properties interesting from the cosmological point of view (their mere enumeration allows one to see to which extent establishing of these properties is the result of both theory and observation.

[1] As it is shown in the Appendix, such a model is in fact not consistent with Newton's theory.

[2] More precisely, these spaces are defined by the action of certain symmetry groups, see, for example, MacCallum 1973.

(1) Observations are carried out at wavelengths 0.1–10 cm. The radiation does not come from any identifiable discrete sources. Most probably, it was emitted "by the Universe" about 10^5–10^6 years after the Big Bang; before that date, called the *epoch*, or sometimes, more geometrically, the *surface, of the last scattering*, photon remained in equilibrium with other components of the primeval plasma. At the last scattering epoch the Universe became transparent for microwave photons; one speaks of photons decoupled from other forms of matter. If this interpretation is correct, the microwave background radiation carries the information about the Universe as regards how it was at the last scattering epoch. In other words, the background radiation allows us to penetrate backward up to 99.999 % of the evolution of the Universe.

(2) The background radiation spectrum has, up to good approximation, a Planck spectrum of the black body at temperature $T = 2.7\,\mathrm{K}$. This means that, at the epoch of emission of the background photons, the Universe was in the thermodynamical equilibrium.

(3) The background radiation possesses a high degree of isotropy. Expected deviations from isotropy can be of a threefold kind:

(a) anisotropy with a 24–hour period (*dipole anisotropy*); it is caused by the Earth motion with respect to the radiation field (once every 24 hours antenna crosses the direction in which the Earth moves with respect to the background radiation field);

(b) anisotropy with a 12–hour period (*quadrupole anisotropy*); it could serve as a test of the anisotropic expansion of the Universe at the last scattering epoch (twice every 12 hours antenna crosses a privileged direction of the world's expansion);

(c) anisotropy on small scales would constitute an empirical proof of small scale inhomogeneities of the Universe at the last scattering surface (the radiation was scattered on inhomogeneities).

The anisotropy with a 24–hour period has indeed been found. It turns out that — after taking into account corrections for the Earth's motion around the Sun and the Sun's motion around the centre of our Galaxy — one must ascribe to our Galaxy a speed of about $540\,\mathrm{km\cdot s^{-1}}$ with respect to the radiation field (in the direction: $\alpha = 10^{\mathrm{h}},7$, $\delta = -22°$). It is worth noticing that the background radiation field is a physical realization of the idea of a co–moving reference frame.

The background radiation field has a high degree of isotropy of types (b) and (c). The upper limit of the temperature T anisotropy, at any scale, is $\Delta T/T < 10^{-3}$, and for angular scales less than 3°, $\Delta T/T < 10^{-5}$. The last result comes as a surprise since, according to present theories of formation of galaxies and cluster of galaxies, pregalactic condensations of matter should manifest themselves in small scale anisotropies of the background radiation. The actually measured degree of the isotropy puts severe constraints on theories of galaxy formation.

To sum up, one must say that observational data concerning the microwave background radiation inform us about: (1) a high degree of its isotropy with respect to the Earthly observer (after taking into account the observer's motion with respect to the co-moving frame); (2) homogeneity of matter distribution on the last scattering

surface (the lack of small scale inhomogeneities); (3) the isotropic expansion at the last scattering epoch (the lack of 12-hour anisotropies). The often overlooked fact is worth noticing; it is that the background photons carry information concerning not only the spatial isotropy of the matter distribution in the Universe, but also its spatial homogeneity (conclusion (2)). (More about cosmological significance of the microwave background radiation, see Peebles 1971, pp. 121–158; Weinberg 1972, pp. 506–228; Sciama 1971, pp. 176–203; Demiański 1985, pp. 274–279; Raine 1981b, pp. 44–61; Bajtlik 1984; Kolb and Turner 1990, pp. 14-15.) Recent observational results and their analysis can be found in the thesis by Bajtlik (1984).)[3]

(B) Investigations of the spatial distribution of galaxies is a difficult observational task. Till not long ago, distances to very few galaxies were known (with large measurement errors); without knowing the "third coordinate" one could only determine the projection of the galaxy distribution onto the celestial sphere. The reconstruction of the three-dimensional distribution from observational data (with no knowledge of the third coordinate) is not unique, and leads only to probable result. From among several methods used to reach the goal, the so-called *correlation function* method turned out to be the most effective one. Two-point correlation function measures the probability of finding (on the celestial sphere) two galaxies such that the distance between them is smaller than the average distance between two neighbouring galaxies. Three-, four-, ..., and n-point correlation functions are defined analogously (for precise definitions, see Peebles 1980, pp. 143–152). Time and effort consuming investigations, carried out with the help of this method, have led to the following results:

(1) For large angular distances the correlation function tends to zero. This means that the distribution of galaxies is homogeneous at large. Observations suggest that the world can be regarded as spatially homogeneous at distance scales larger than 10^8 ps.[4]

(2) On smaller scales a strong effect occurs of the clustering of galaxies. This effect remains in agreement with the well-established existence of smaller aggregations of galaxies (such as the Local Group), clusters, and superclusters of galaxies.

(3) The investigations do not show any characteristic scale of clustering. This means that there is no privileged unit (cluster of galaxies, say) which could be regarded as a fundamental building-block of the Universe. The average distribution of such building blocks would implement the observed large scale distribution of matter. This suggests the hypothesis that there is clustering of galaxies on all scales: smaller aggregates tend to be clustered into larger ones, but this tendency is less and less pronounced as one goes to larger and larger distances.

[3]The above presented results have been remarkably improved by the very successful satellite COBE (COsmic Background Explorer) launched in November 1989. Preliminary measurements fit perfectly to a Planck blackbody radiation spectrum for a temperature of 2.735 ± 0.06 K. At the 1% level there are no deviations from an ideal blackbody spectrum. Dipole and quadrupole anisotropies are not observed. See B. Schwarzschild, *Physics Today*, **43**, 1990, 17–20. (Note added in proof.)

[4]1 ps = 206 265 AU = 3.263 light years = 3.083×10^{13} km (AU stands for Astronomical Unit).

Many red shift measurements of galactic spectra have been recently obtained (see below (C)). By assuming that the red shifts provide information on distances to galaxies, the lacking third coordinate can be determined for a considerable number of galaxies, and three–dimensional charts of the sky can be constructed. This kind of research has confirmed the previous results but, quite unexpectedly, revealed a peculiar structure of clustering on smaller scales. It turns out that galaxies are clustering in elongated membrane structures the linear size of which is of order of magnitude larger than the transversal one. Between membranes crossing each other in an irregular manner, large domains extend devoid of galaxies. Within the membranes galaxies are distributed homogeneously with a tendency of clustering on all scales. (See Silk et al. 1983; more about distribution of galaxies in the Universe, Peebles 1980; Groth et al. 1977; Raine 1971b, pp. 1–33, 62–83; Kolb and Turner 1990, pp. 16–26.)

Important cosmological information can be obtained by comparing empirical results concerning the microwave background radiation and the distribution of galaxies. In this way, one can get, as it were, two photographs of the Universe: the background radiation conveys information about the structure of the last scattering surface, the distribution of galaxies–about the "present" cosmological epoch. It turns out that the evolution leads the Universe towards the destruction of homogeneity. This effect should be ascribed to a strongly unstable character of gravitational field which consists in a tendency to strengthen all casual matter inhomogeneities. Onto this general trend other physical mechanisms are superimposed the nature of which has not been fully understood. The origin and evolution of structures in the Universe is one of the most difficult problems of contemporary cosmology.

(C) Red shifts of the spectra of galaxies were the first significant observable of the relativistic cosmology, it played an exceptional role in its history. Now, we will show how red shifts are connected with theoretical concepts appearing in the standard model.

Let us consider Robertson–Walker line element (3.1)–(3.2b). Let us assume that at an instant t_1 of the cosmological time a photon is emitted from a galaxy situated at the point $(r_1, \theta_1, \varphi_1)$ in the direction r (i.e., $\theta = \varphi = 0$). The photon is registered on the Earth, situated at $(0,0,0)$, at the instant t_0. Taking into account that light moves along null geodesics, one obtains

$$\int_{t_1}^{t_0} \frac{dt}{R(t)} = \int_0^{r_1} \frac{dr}{1 + \frac{1}{4}r^2} \equiv f(r_1), \qquad t_0 > t_1. \tag{6.1}$$

Since the coordinates $(r_1, \theta_1, \varphi_1)$ of the galaxy are co–moving coordinates, $f(r_1)$ does not depend on time; it can be thought of as measuring a conventional distance corresponding to the time difference $t_0 - t_1$ (conventional since it is not directly measurable). If the subsequent photon is emitted from the galaxy at an instant $t_1 + \Delta t_1$ and registered on the Earth at $t_0 + \Delta t_0$ with Δt_1, $\Delta t_0 \ll t_0 - t_1$, one has

$$\int_{t_1+\Delta t_1}^{t_0+\Delta t_0} \frac{dt}{R(t)} = f(r_1),$$

or

$$\frac{\Delta t_0}{R(t_0)} = \frac{\Delta t_1}{R(t_1)}.$$

The last formula is often expressed in terms of the wavelength λ_1 (or frequency ν_1) of the emitted light and of the wavelength λ_0 (or frequency ν_0) of the registered light; in such a case we have

$$\frac{\lambda_1}{\lambda_0} = \frac{\nu_0}{\nu_1} = \frac{\Delta t_1}{\Delta t_0} = \frac{R(t_1)}{R(t_0)}.$$

Customarily, one defines the *red shift parameter* (called also *red shift*, for short) $z = (\lambda_0 - \lambda_1)/\lambda_1$ which gives

$$z = \frac{R(t_0)}{R(t_1)} - 1. \tag{6.2}$$

If the Universe expands, $R(t_0) > R(t_1)$ and z is positive, i.e., the registered light wave is shorter than the emitted one. If the Universe contracts, the situation is just the opposite, in which case one obtains blue shift. By expanding (6.1) and (6.2) with respect to the parameter $h = (t_0 - t_1)/R(t_0)$, which is assumed to be small (i.e. one could hope that the procedure is valid for not too distant galaxies), and then by eliminating this parameter from expansions of (6.1) and (6.2), one obtains

$$z = \dot{R}(t_0) f(r_1).$$

This gives a linear dependence of red shift on the "distance" $f(r_1)$. Distances, actually measured by astronomers (see Weinberg 1972, pp. 418–427) turn out to be proportional to $R(t_0)f(r_1)$, symbolically $d \cong R(t_0)f(r_1)$, i.e.,

$$z \cong \frac{\dot{R}(t_0)}{R(t_0)} d. \tag{6.3}$$

This is the famous *Hubble's law*, discovered observationally by this astronomer in 1929. It was this law that, before the discovery of the background radiation, constituted the main argument of the "expanding Universe".

Before the discovery of the microwave background radiation the so-called *classical cosmological tests* were expected to provide relevant observational information. The idea consists in identifying actually measured astronomical quantities (sometimes called *observables*) with theoretical magnitudes appearing in a model, and comparing dependencies between pairs of them, as predicted by various models, with those established experimentally. The most important dependencies are: red shift — angular size of objects, red shift — stellar magnitude, counting of objects up to a certain limiting magnitude[5] (see for instance, Weinberg 1972, pp. 441-458; Raine 1981b, pp. 153-166). However, so far the efficiency of these tests has proved to be rather limited.

[5]In cosmological tests magnitude often serves as a distance indicator.

They are based on a number of in principle unverifiable assumptions which have to fill in gaps in our knowledge, especially as far as the evolution of objects is concerned (such as galaxies, radiogalaxies, quasars). By maneuvering "free parameters" many different cosmological models can be fitted to actually observed dependencies. The most significant result obtained with the help of classical tests is the support given by them to the idea of the evolutionary Universe against the steady–state cosmology of Bondi and Gold (1948), and Hoyle (1948).

To sum up, one must say that the standard world model presented in the present chapter does not contradict any observational data available at present and, although none of known tests "proves" it directly, all tests together strongly suggest that the standard model — up to a good approximation — correctly describes the observed Universe.

We should notice another important circumstance as far as observational aspects of cosmology are concerned. Because of the final velocity of light signals, the so–called *horizon* may appear in the Universe limiting the field of vision of a given observer. At a moment t, an observer situated at the beginning of the co–moving coordinate system, $r = 0$, can receive signals emitted by galaxies (more strictly by particles at rest with respect to the co–moving coordinate system), the radial coordinate of which, r, is smaller than r_1, where r_1 is the coordinate of a galaxy such that the light signal emitted from it at t_1 reaches the observer (situated at $r = 0$) at t. t_1 can be computed from the formula (see formula (6.1))

$$\int_0^{r_1} \frac{dr}{1 + \frac{1}{4}r^2} = \int_{t_1}^{t_0} \frac{dt}{R(t)} \tag{6.4}$$

If the integral on the right–hand side of (6.4) converges for $t_1 \to 0$ (or in world models without the initial singularity, for $t_1 \to -\infty$), then there exist galaxies such that light signals emitted from them do not have enough time to reach the observer at $r = 0$; such a model is said to have a *particle horizon*.

It can also happen that the integral on the right–hand side of formula (6.4) converges for $t \to t_{\max}$ ($t_{\max}$ can be equal to $+\infty$ or to the moment of the final singularity); if this is the case, there exist galaxies such that light signals emitted from them will never reach the observer situated at $r = 0$; such a model is said to have an *event horizon*. (More about horizons, see Rindler 1946; 1969, pp. 240–242; Weinberg 1972, pp. 489–491; Hawking and Ellis 1973, pp. 127–130.)

The existence of horizons imposes strong constraints upon observational possibilities of cosmology. If in our Universe there are horizons, and if we want to extrapolate our knowledge acquired within the horizons to regions situated outside them, we have to make strong assumptions that would justify such a procedure. Assumptions of this kind will always be of an empirically unverifiable character.

7. *Comments and Remarks*

A. After having made the first contact with the standard cosmological model, one is struck by a certain tension between the degree of idealization incorporated into this model and the agreement of its predictions with observational data. Definitions of particle, photon, etc. from the beginning of sec. 2 allow one to "theoretically identify" the material content of the standard model; however, to coordinate physical objects to particular mathematical structures becomes a very difficult task. For instance, in the standard model, density of particles with non–zero rest–mass satisfies the relationship $\rho_{\text{mat}} \sim R^{-3}(t)$, whereas for particles with zero rest–mass one has $\rho_{\text{rad}} \sim R^{-4}(t)$ (number of photons in the volume unit changes as R^{-3}, and the energy of each photon as R^{-1}). If, therefore, in the present Universe, there is even a slight amount of radiation, there had to be in its past an epoch (when R was sufficiently small) in which the radiation dominated the dynamics of the world. If the present Universe is correctly described by a dust–filled model with even a slight admixture of radiation, its sufficiently distant past must be described by a radiation–filled model.

In its "dust period" the model is filled with dust particles, i.e., with timelike geodesics (definition 2.2), but it is less clear with which physical objects a "theoretical dust particle" should be identified. From the theoretical point of view, dust particles are distributed throughout the space of the model in a homogeneous and isotropic manner, and consequently one must look for real objects on the level of which the spatial homogeneity and isotropy would be guaranteed. Galaxies for sure, and clusters of galaxies most probably, are not such objects. If a version of the hypothesis is true of clustering of galaxies on all scales, it might turn out that such objects simply do not exist. Having this in mind, successes of the standard world model in organizing observational material are really astonishing. (For a more detailed discussion, see Heller 1982.)

B. All solutions of Einstein's equations, considered in the present chapter, have been obtained with the assumption that the cosmological constant Λ is equal to zero (see chapter 4, sec. 5). By rejecting this assumption one obtains a much richer class of solutions. Counterparts of equations (3.3) and (3.4) with non vanishing cosmological constant are the following,

$$\frac{8\pi}{3}\rho = \frac{\dot{R}^2}{R^2} + \frac{k}{R^2} - \frac{1}{3}\Lambda,$$

$$-8\pi p = 2\frac{\ddot{R}}{R} + \frac{\dot{R}^2}{R^2} + \frac{k}{R^2} - \Lambda.$$

Solutions of these equations, for different values of the parameters Λ and k, with the equation of state $p = 0$ (dust), are shown in fig. 6.4 and 6.5; the table gives their straightforward classification. (For more details, see Tolman 1934, pp. 394–419; Rindler 1977, pp. 235–238; Weinberg 1972, pp. 613–616.)

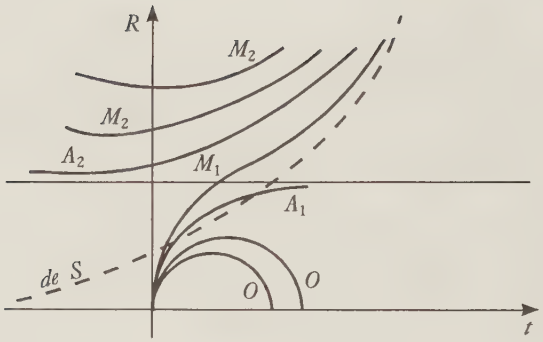


Fig. 6.4. World models with $k = +1$, $p = 0$, for different values of the cosmological constant.

Einstein (1917) introduced the cosmological constant Λ to obtain a static world model; however, after finally convincing himself that the Universe really expands, he considered it to be superfluous (see Einstein 1958). Observational data testify that the influence of the cosmological constant upon the dynamic of the planetary system is negligible. The interest in cosmological models with the cosmological constant increased around 1967 when a surplus of number was discovered of quasars with red shifts about $z = 2$. It seemed that this suggested the existence of an epoch in which the rate of expansion slowed down almost to zero. Such an epoch appears only if the value of the cosmological constant Λ is slightly larger than Λ_E (the so–called *Lemaître's models*, see model $\mathcal{M}_1$ in fig. 6.4). When afterwards it had turned out that it was only a kind of a selection effect, models with $\Lambda = 0$ became again fashionable (see Petrosian 1974).

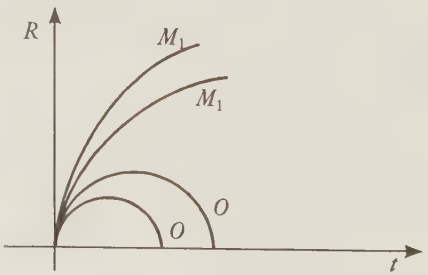


Fig. 6.5. World models with $k = 0$, $p = 0$, for different values of the cosmological constant.

Table. World models with the cosmological constant.

$\Lambda \backslash k$		$= +1$	$= 0$	$= -1$
> 0	$> \Lambda_E$	M_1 – monotonically expanding models with the initial singularity. $M_1 \to$ de S as $t \to \infty$	Models M_1	Models M_2
	$= \Lambda_E$	$R < R_E$: A_1 — first asymptotic model. $A_1 \to E$ as $t \to \infty$		
		$R = R_E$: E — static Einstein's model.		
		$R > R_E$: A_2 — second asymptotic model. $A_2 \to E$ as $t \to \infty$		
	$< \Lambda_E$	$R < R_E$: O — oscillating models with the initial and final singularities		
		$R > R_E$: M_2 — monotonically expanding models without the initial singularity. $M_2 \to$ de S as $t \to \infty$		
$= 0$	$= 0$	Models O		
< 0	< 0		Models O	Models O

de S denotes de Sitter's world model; Λ_E and R_E are values of the cosmological constant and of the scale factor characteristic for the static Einstein's model, correspondingly.

The cosmological constant has returned to scientific papers in connection with grand unification theories. It was known for a long time that dynamical effects of the cosmological constant appear as "corrections" to the density and pressure: $\rho_{ef} = \rho + \Lambda/(8\pi)$, $p_{ef} = p - \Lambda/(8\pi)$ (see Weinberg 1972, p. 614).[6] The energy density of the vacuum presupposed by unifying theories is thought of as being connected with these corrections (see, for instance, Kazanas 1980, Abbot 1988).

C. Einstein's equations, being partial differential equations, determine the structure of space–time only locally. In particular, they do not contain any global information about space–time topology. For instance, for a given space time $(\mathcal{M}, g)$, there exist a simply connected universal covering manifold[7] $\tilde{\mathcal{M}}$ equipped with the metric $\tilde{g}$ such that $(\mathcal{M}, g)$ can be obtained from $(\tilde{\mathcal{M}}, \tilde{g})$ by suitably identifying points in $\tilde{\mathcal{M}}$. If a solution of Einstein's equations is given, it is always possible to find its universal covering $(\tilde{\mathcal{M}}, \tilde{g})$, and to perform all possible identifications in it (if space time has no symmetries its universal covering coincides with it). Space-times obtained in this

[6] To obtain a static solution from the field equation without the cosmological constants requires a negative pressure. The cosmological constant was introduced by Einstein in order to avoid this "non–physical" effect.

[7] Definitions of these concept can be found, for example, in a book by O'Neill (1983, pp. 441–445).

way are automatically solutions of Einstein's equations, but they can drastically differ from each other from the topological point of view. In sec. 3 of the present chapter, it has been said that the standard choice of topology for instantaneous spaces are hyperbolic space $\mathcal{H}^3$ for $k = -1$, Euclidean space $\mathbf{R}^3$ for $k = 0$, and spherical space $\mathcal{S}^3$ for $k = +1$. However, the problem could be stated in a more general way, namely, one could ask about all possible space–times which locally are expanding Robertson–Walker worlds with instantaneous spaces of constant curvature. The answer to this question gives globally different cosmological models that locally are non–distinguishable from each other (and from the standard model) as far as their observational aspects are concerned. Global differences between these models could manifest themselves in the possibility of observing the same object more than once on the celestial sphere (the so–called *ghost images*).

Mathematically, the question is reduced to the classical problem of *Clifford–Klein space forms* (see Wolf 1967, and in a more concise way O'Neill 1983, pp. 243–248). It was Ellis (1971) who considered this problem in a cosmological context. It turns out that there exist infinitely many space forms[8] locally reproducing Robertson–Walker universes. In the case $k = +1$, all these space forms are compact but in the cases $k = 0$ and $k = -1$, there exist both compact and non–compact space forms. This means that not all world models with $k = 0$ and $k = -1$ should be regarded as "open".

This problem is strictly connected with the extrapolation problem in cosmology which essentially reduces itself to the question: how spaces, determined locally, could be extended to non–local scales (see Heller 1978)? Some observational aspects of this problem have been discussed by Ellis (1983).

D. The existence of horizons in some world models (see end of sec. 6) also leads to difficult cosmological problems. Let us notice that two domains of space–time separated by a particle horizon never have interacted with each other. It can be shown that domains, distant from each other by no more than 75° on the celestial sphere, were separated by the particle horizon at the epoch at which the microwave background radiation interacted for the last time with other forms of matter filling the Universe (see, for example, Weinberg 1972, pp. 525–526). If one admits that physical interactions smoothing out the Universe cannot propagate with a speed greater than that of light, one cannot see any reason why the background radiation, coming from domains separated from each other by a particle horizon during the last scattering epoch, is now isotropic to such a high degree.

If one postpones looking for a physical mechanism explaining this puzzle, one has to assume very special initial conditions for the Universe which would *a priori* synchronize cosmological parameters (such as temperature of the background radiation) in regions separated by particle horizons from the very beginning.

A program aiming at explaining the question of the "smoothness of the Universe"

[8] A complete and connected Riemann manifold of constant curvature is said to be a (*Clifford–Klein*) *space form*.

was proposed by Misner (1969). The idea was to find such mechanisms with the help of which it would be possible "to deduce" the actually observed state of the Universe from *arbitrary* initial conditions (this program has been named the *mix-master cosmology*). To this end one must (1) find world models in which a light signal would be able to go many times around the world till the present epoch, and (2) identify physical mechanisms which could effectively smooth out the Universe. Different mechanisms were proposed: among others dissipative effects connected with the viscosity of neutrinos or with the quantum process of the pair particle–antiparticle creation in strong gravitational fields in the early stages of world's evolution. The heated discussion, which developed around the mix–master cosmology program, has shown that to satisfy condition (1) very special initial data are also required; therefore, it is not a good strategy to avoid special conditions. As far as condition (2) is concerned, it has turned out that the proposed mechanisms are either not efficient enough (neutrino viscosity) or too hypothetical (pair creation). In effect, the mix-master program is believed to be attractive but involved in so far insurmountable difficulties (see Barrow and Matzner 1977, Barrow 1978, Misner 1980, Raine 1981b, pp. 215–220).

New perspectives to solve the puzzle appeared in connection with the grand unification theories. Guth (1981) noticed that the phase transition from the early symmetric Universe to the Universe with the spontaneously broken symmetry can be associated with a short period of rapid (exponential) expansion enlarging the size of the Universe 10^{50} times (or more). "Inflation" of this kind is caused by the vacuum energy which, during the phase transition, dominates the cosmological dynamics. This scenario, called the *inflationary universe*, was afterward modified several times (see Linde 1982, Hawking and Moss 1982, Barrow and Turner 1982). Inflation seems to liquidate the horizon problem: because of the exponential expansion in its history, the Universe, observed at present, was never divided by particle horizons; it had originated from a tiny portion of the primordial plasma blown up to enormous size during the inflation period.

Great hopes were connected with inflationary scenarios. However, the enthusiasm which surrounded early versions of the model has recently slowed down. Some researches point out that physics, presupposed by inflation, is still dubious, and for solutions of some problems one is paid with new, and even more serious, problems (see Ellis and Stoeger 1988).

Bibliographical Essay

Reading of the present book does not give an idea of what the contemporary cosmology really is. The book's aim has been only to pose the cosmological problem and to discuss its fundamental assumptions. If the reader feels a little disappointed that he did not acquire the expected cosmological knowledge, the author is entitled to think that he has succeeded to reach his goals. However, wanting to help the reader in his further adventures with cosmology, the author feels obliged to add to this book a kind of guide throught the land of cosmological readings. It should be stressed that it is by no means a bibliography which could pretend to even partial completeness. The term "guide" seems to be a just expression; it is a guide, and — it should be added — a subjective one, to a great extent reflecting pathways travelled once by the author himself.

The guide contains only references to the literature in the field of relativistic cosmology; books dealing with other cosmologies (such as neo–Newtonian, Milne's or steady–state cosmology) are not included. With only one exception (Robertson's review article of 1933) the reader will not find in it any reference to scientific papers published in journals. Because of the great richness and variety of such papers any attempt to compose a guide through them, or a list of them, would be both groundless and impossible.

For reasons of clarity, cosmological literature will be divided into sections reflecting main stages of the evolution of cosmology in our century.

A. Classical Period (from the origin of relativistic cosmology to the Second World War).

First textbooks of general relativity appeared surprisingly early. To the earliest ones belong:

H. Weyl (1922), *Space — Time — Matter* (German edition in 1921);

A.S. Eddington (1923), *The Mathematical Theory of Relativity.*

As far as mathematical standards are concerned, both these books are equal to many later publications, and the depth of their philosophical insight surpasses many newest elaborations. Numerous comments in these books are devoted to the problem of the relativity of space and time; the problem was still new and claimed for elucidation. Only some sections deal with "the Universe as a whole"; they represent Einsteins's "cylindrical" world and de Sitter's "spherical" universe. The fact that the two world models exist fascinates the authors. For the first time space–time begins to reveal its global image.

About 1930, owing mainly to the works of Friedman, Lemaître and Robertson, all spatially homogeneous and isotropic cosmological models were known. The first complete review of them can be found in the paper:

H.P. Robertson (1933), *Relativistic Cosmology*,
published in the journal *Reviews of Modern Physics*. The article is important for two reasons: first, a few generations of physicists learned from it basic facts of relativistic cosmology; second, appendix to the paper contains an almost complete list of cosmological papers published in 1917–1932.

An extensive monographic textbook:

R.C. Tolman (1934), *Relativity, Thermodynamics and Cosmology*,
contains a detailed presentation of special and general relativity; in this aspect it is certainly a kind of textbook. Chapters that present pioneering works of Tolman himself are doubtlessly of a monographic character; they create the fundamentals of relativistic thermodynamics. The cosmological part of the book, in its composition, is similar to the paper by Robertson of 1933 but it puts more stress on the confrontation of theoretical models with astronomical observations. Applications of relativistic thermodynamics to cosmology are typical for this book. The clarity of Tolman's work is exceptional. Even today scientists open it to check a formula or to find a forgotten, but still useful, mathematical expression.

One should also mention another book which — although published after the Second World War — by its spirit belongs to the pre–war period:

G. Lemaître (1972), *L'Hypothese de l'atome primitif* (first edition in 1946; English translation in 1950).

The book is a collection of popular essays, published earlier by the author on various occasions. The value of the book lies in the fact that it shows the whole cosmological vision of Lemaître. The hypothesis of the primeval atom was a first attempt to create an overwhelming cosmological system (from the initial singularity, through a cosmological theory of nucleosynthesis, to a theory of the origin of galaxies and of clusters of galaxies). In this hypothesis, for the first time, geometric theory of the world was supplemented with a reconstruction of physics of the early Universe.

The book:

O. Heckmann (1968), *Theorien der Kosmologie* (first edition in 1942)
closes this period. In the exposition style it follows Robertson's paper (1933). It also contains a list of cosmological publications of 1933–1940.

B. Period of Stagnation (from the Second World War to the discovery of the background radiation in 1965).

First an explanation. The term "stagnation" implies a certain slowing down of progress in cosmology, but it does not apply to the theory of relativity which had many successes in this period especially as far as its mathematical aspects were concerned. Of course, all these successes were relevant also for cosmology. Speaking about stagnation I have in mind a certain impasse in cosmological observations. In the fiftieth discussion focused on polemics between defenders of the relativistic paradigm

(which was then called the "theory of the expanding Universe") with the proponents of the steady–state cosmology, created in 1948 by Bondi and Gold, and Hoyle.

A note of pessimism can be heard as early as in 1945 in the Cosmological Appendix added by Einstein to the second edition of his *The Meaning of Relativity* (first edition in 1921). A small number of publications in the field of cosmology that are to be quoted in the present section also testifies about the stagnation. We should certainly mention

H. Bondi (1960), *Cosmology* (editions in 1952, 1960, 1961).

The book presents "on equal footing": neo–Newtonian cosmology, relativistic cosmology, Milne's cosmology, and steady–state cosmology. Cosmological conceptions of Eddington, Dirac, and Jordan are also briefly reported. A transparent discussion of the conceptual and methodological basis of cosmology makes this book still currently relevant.

The following book, on the one hand, is a kind of summing up of formal achievements of the pre–war cosmology but, on the other hand, contains a number of theoretical results obtained already in this period:

H.P. Robertson and T. W. Noonan (1968), *Relativity and Cosmology.*

In spite of the date of its publication, this book doubtlessly belongs to the period before the discovery of the background radiation. After the tragic death of Robertson (in a traffic accident), it was compiled, from papers and notes left by him, by his last student, Thomas W. Noonan. The book was very carefully edited;[1] it shows the Robertsonian style of doing cosmology, in which a detailed analysis of space–time symmetries is as important as solving the field equations.

C. Period of Dynamical Progress (after the discovery of the background radiation).

The discovery of the microwave background radiation in 1965 together with other achievements of radioastronomy (quasars, radiogalaxies, pulsars) and later on also of optical astronomy brought new momentum into relativistic physics and especially into cosmology. Relatively quickly, the standard model began to consolidate. From the beginning of the seventieth the number of books in cosmology continuously increases. Only some of them can be quoted. Because of their great variety, I shall classify them into certain groups.

1. *Textbooks of general relativity.* It is now a custom to devote a chapter or (more often as time passes) an entire part to cosmology. A handbook which has gained considerable publicity is

C.W. Misner, K. Thorn, J.A. Wheeler (1973), *Gravitation.*

[1] At the end of the book the reader can find summaries of all chapters. The summary ends with the summary of the summary.

It is a big volume (1279 pages). The style is highly didactic with the ambition to teach a beginner all the mysteries and intricacies indispensable to start research in this field. (In my opinion, didactic goals of this book are a bit exaggerated.) Many pages are devoted to cosmology, but a serious drawback (in my opinion again) is that the authors, motivated by philosophical reasons, almost exclusively focused on "closed" world models (with $k = +1$).

The more recent book:

B. Schutz, *A First Course in General Relativity*

bears some similarities to Misner, Thorn and Wheeler's volume, but is much more concise, easier, and with no tricky gadgets. It introduces the reader to understanding the content of the theory presented and to manipulating with its computational machinery. This book is my personal recommendation for an ambitious freshman.

From among many other handbooks I would recommend the following to a reader who is prepared to make a more intimate acquaintance with cosmology:

S. Weinberg (1972), *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity.*

The fifth part of this book deals with cosmology. In my view, only two features separate the book from the ideal: first, the author unnecessarily avoids the modern coordination free notation; second, the book was written several years ago, and in a few points it would require modernization. Nevertheless, the principles of physics do not change so rapidly and Weinberg's book is still an excellent source to learn relativistic physics and cosmology.

Let us also quote:

W. Rindler (1968, 1977), *Essential Relativity: Special, General, and Cosmological.*

R. Adler, M. Bazin, M. Schiffer (1965), *Introduction to General Relativity.*

The first of these books, providing more imaginative pictures, focuses on the problem of the relativity of space and time (Mach's principle), and from the second the reader can learn many useful mathematical techniques.

L.D. Landau, E.M. Liphschitz (1973), *The Field Theory* (in Russian).

This edition of the famous book has a significantly enlarged cosmological part as compared with previous editions. A serious drawback, from the cosmological point of view, is the absence of observational aspects of cosmology. A chapter devoted to cosmology can be found in the book

M. Demiański (1985), *Relativistic Astrophysics.*

The following book is of an exceptional character,

R.S. Sachs and H. Wu (1977b), *General Relativity for Mathematicians.*

It adopts a quasi mathematical style of presentation. However, mathematical rigors break down when a physical reality has to be put into formulae. In the part dealing with cosmology, the Einstein–de Sitter world model is mainly exploited (it is a spatially isotropic and homogeneous world model with $k = p = \Lambda = 0$).

Very useful for every relativity student is the collection of exercises:

A.P. Lightman, W.H. Press, R.H. Price and S.A. Teukolsky (1975), *Problem Book in Relativity and Gravitation.*

Chapter 19 is on cosmology (solutions to problems are included).

2. *Handbooks of cosmology.* a) Elementary level:

(1) *Understanding Space and Time*, Block 5: *Cosmology* (1979),

(2) M. Berry (1976), *Principles of Cosmology and Gravitation*,

(3) M. Rowan–Robinson (1977), *Cosmology*,

(4) D.W. Sciama (1975), *Modern Cosmology*,

(5) J. Heidmann (1973), *Introduction à la cosmologie*,

(6) P.T. Landsberg, D.A. Evans (1977), *Mathematical Cosmology — An Introduction.*

The above books are enumerated, roughly speaking, in the order from the easier to the more difficult. I would advise one to make the first encounter with cosmology with the help of (1) or (2). (1) is a publication of the British "Open University", (2) is an excellent textbook for beginners. (3) deals more with the observational side of cosmology, and (4) nicely presents mutual dependencies of cosmology and astrophysics. (5) is a sample of the French approach to cosmology. By using (6) one can acquire a proficiency in "mathematical cosmology" without going beyond the fundamental level of higher mathematics. Although the book deals, in principle, with the neo–Newtonian cosmology only, it gives a good picture of cosmological research.

b) Advanced level:

(1) D.J. Raine (1981b), *The Isotropic Universe*,

(2) P.J.E. Peebles (1971), *Physical Cosmology*,

(3) J.B. Zeldovitch and I.D. Novikov (1975), *Structure and Evolution of the Universe* (in Russian).

(4) A.D. Dolgov, I.B. Zeldovitch and M.B. Sagin (1988), *Cosmology of the Early Universe* (in Russian).

(5) M.P. Ryan and L.C. Shepley (1975), *Homogeneous Relativistic Cosmologies.*

These books are also enumerated in the order of increasing difficulty. (1) is a good introduction to, and overview of, cosmology. (2) and (3) are attempts at reconstructing physical processes in the evolving Universe. (3) is a mosaic of various cosmological problems rather than a textbook of cosmology, albeit a mosaic providing the general picture. (4) originated from courses given by Zeldovitch at the Moscow University, and was composed by his co–workers after the death of this eminent cosmologist; it gives a very penetrating account of recent problems and results in cosmology, such as: inflation, grand unification, quantum gravity, creation of the Universe as a quantum tunneling process, ... (5) is mathematically elegant and explores theoretical aspects of Bianchi cosmology (homogeneous but not necessarily isotropic world models). This book contains a rich catalogue of cosmological publications till 1974, which can be considered as a continuation of similar lists composed by Robertson (1933) and Heckmann (1968) (however, now without any ambition of completeness).

3. Monographs.

(1) P.J.E. Peebles (1980), *The Large Scale Structure of the Universe*,

(2) S.W. Hawking, G.F.R. Ellis (1973), *The Large-Scale Structure of Space-Time*,

(3) J.K. Beem, P.E. Ehrlich (1981), *Global Lorentzian Geometry*,

(4) D. Kramer, H. Stephani, E. Herlt, M. MacCallum (1980), *Exact Solutions of Einstein's Field Equations*.

(1) deals with the problem of the origin and evolution of structures in the Universe (such as galaxies, clusters of galaxies, ...). Works which gave birth to book (2) have initiated many new research programs in relativity: singularity theorems, causal structure of space-time, global structure of solutions to Einstein's field equations, and so on. This monograph closes a certain stage of discussions concerning the classical (non-quantum) singularity problem. (3) elaborates principally the same problems as (2) but in a more mathematical style. (4) is an indispensable source for studying solutions of Einstein's equations.

4. Books on the XX^{th} century history of cosmology.

(1) J.D. North (1965), *The Measure of the Universe*,

(2) J. Merleau-Ponty (1965), *Cosmologie du XX^{e} siecle*,

(3) O. Godart and M. Heller (1985), *Cosmology of Lemaître*.

(4) E.A. Tropp, W.Ya. Frenkel, A.D. Chernin (1988), *Alexander Alexandrovitch Friedman* (in Russian).

5. Books on philosophical aspects of relativity and cosmology.

(1) J.D. Barrow and F.J. Tipler (1985), *The Anthropic Cosmological Principle*,

(2) D.J. Raine and M. Heller (1981), *The Science of Space-Time*,

(3) R. Toretti (1983), *Relativity and Geometry*,

(4) M. Friedman (1983), *Foundations of Space-Time Theories*,

(5) J.D. Barrow (1988), *The World Within the World*,

(6) M. Heller (1986), *Questions to the Universe*.

6. Popular books on cosmology. Since a great number of such books appears each year, I choose, more or less at random, only a few:

(1) J. Silk (1980), *The Big Bang. Creation and Evolution of the Universe*,

(2) E.R. Harrison (1981), *Cosmology – The Science of the Universe*,

(3) S. Weinberg (1977), *The First Three Minutes*,

(4) R. Geroch (1978), *General Relativity from A to B*,

(5) A.G. Pacholczyk (1984), *The Catastrophic Universe*,

(6) J. Silk and J.D. Barrow (1983), *The Left Hand of Creation*,

(7) J.S. Trefil (1983), *The Moment of Creation*.

My personal recommendation would be (1), (6) and (7).

Appendix

Evolution of Dynamical Space–Time Theories

0. Introduction

General relativity is something more in contemporary physics than one of the many theories describing a certain domain of the world. A distinguished role of Einstein's theory of gravity can be reduced to two circumstances: first, this theory provides a space–time stage to many other physical theories and, second, it does so not only *locally*, in the neighbourhood of an investigated event, but also globally when, for instance, it deals with inextendible space–times, i.e., with such space–times which cannot be embedded into "larger" ones. Doubtlessly, the last function of general relativity appeals to aesthetic feelings of physicists; however, its significance in our system of knowledge is both more prosaic and more important: without it the correct statement of many concrete problems of physics and astronomy would be impossible. Contemporary observational astronomy controls such extensive regions of the Universe that with no help of global (or at least large scale) theory of space–time there could be no question of distilling any information carried by signals travelling through space and time. We have also seen in previous chapters how cosmology becomes unavoidable as far as the program of unification of physics is concerned (see chapter 3, sec. 0).

Science always tended towards a synthesis of its various theories. In previous paradigms, there were dynamical theories of mechanics that played the role of preparing spatio–temporal stage for eventual cosmological synthesis. This was the case with both Aristotelian and Newtonian images of the world. In contemporary science this function has been taken over, and much sharpened, by general relativity.

After having read all previous chapters, only a slight mathematical preparation would be enough in order to see, in a more detailed way, how the evolution of views on the space–time structure has led to the present distinguished position of general relativity in science. This could serve us as an additional argument on behalf of this theory — the argument from a "logic of development" of science. Since, on the one hand, analyses of this kind do not fit the main line of reasoning adopted in this book and, on the other hand, it would be a great negligence to omit an occasion to present the place of general relativity in the evolution of our views on the Universe altogether, we will do this in the present Appendix.

To do so, we must prepare certain concepts from the mathematical theory of categories. This theory allows one to organise large parts of contemporary mathematics in the spirit similar to that of the Erlangen Program proposed by F. Klein as a framework for unifying different geometric theories. Some mathematicians share the opinion that this announces a revolution in mathematics analogous to that initiated

by Cantor's work (see Semadeni and Wiweger 1978, p. 13). No wonder, therefore, that the application of the categorial style of thinking to the analysis of physical theories not only turns out to be an attractive possibility, but also begins to bring fruits in the form of significant sharpening of some concepts and of revealing logical dependence between present physical theories and their predecessors (see Trautman 1973b).

Finally, one should stress that analyses carried out in the present appendix are of no strict historical character; they are rather a stylization of the history of science: they try to look at former theories in the light, and with the help, of modern mathematical tools. Such an approach enables one better to see a logic of the evolution of science, and consequently, more responsibly to predict perspectives of further development.

Such analyses have been carried out by many authors (see bibliographical note at the end of the appendix). Of course, we shall make ample use of them (especially of our own study, see Raine and Heller 1981); however, we are in a better position than all previous authors inasmuch as we can base our analysis upon an advanced mathematical machinery prepared throughout the entire book. Owing to this, we are able to significantly shorten the "explanatory text" and provide a deeper insight into the nature of analysed problems.

After presenting elementary concepts of the mathematical category theory (in sec. 1), we shall present the structure and dynamics of space–time in the Aristotelian theory (sec. 2), in classical mechanics without gravitational field (sec. 3), and in classical mechanics with gravitational field (sec. 4). The evolution of these space–time structures naturally leads to special and general relativity theories (sec. 5). Some branching from the main evolutionary trunk, namely certain ideas coming from Berkeley, Leibniz and Mach, together with prospects for future progress, will be presented as *Comments and Remarks* closing the Appendix (sec. 6).

1. Elementary Repetition of the Category Theory

1.1. Definition. We say that a *category A* is given if there are

(1) a class A^0 (not necessarily a set[1]), elements of which are called *objects*;

(2) a function which to every ordered pair of objects $\langle \mathcal{X}, \mathcal{Y} \rangle$, $\mathcal{X}, \mathcal{Y} \in A^0$ ascribes a set $\mathrm{Mor}(\mathcal{X}, \mathcal{Y})$, elements of which are called *morphisms* from $\mathcal{X}$ to $\mathcal{Y}$; additionally

(3) a *composition* of morphism

$$\mathrm{Mor}(\mathcal{X}, \mathcal{Y}) \times \mathrm{Mor}(\mathcal{Y}, \mathcal{V}) \longrightarrow \mathrm{Mor}(\mathcal{X}, \mathcal{V})$$

is defined by

$$(f, g) \longmapsto h = g \circ f,$$

where $f \in \mathrm{Mor}(\mathcal{X}, \mathcal{Y})$, $g \in \mathrm{Mor}(\mathcal{Y}, \mathcal{V})$, $h \in \mathrm{Mor}(\mathcal{X}, \mathcal{V})$; the composition of morphisms is assumed to satisfy the following conditions,

[1] In the sense of the axiomatization of the set theory proposed by Gödel (see, Semadeni and Wiweger 1978, pp. 21–22).

(a) *associativity*: $j \circ (g \circ f) = (j \circ g) \circ f$, where $j \in \mathrm{Mor}(\mathcal{V}, \mathcal{U})$, and f and g are as above;

(b) the existence of *morphism identity*: for every object $\mathcal{X} \in A^0$ there exists one and only one morphism $\mathrm{Mor}(\mathcal{X}, \mathcal{X})$, denoted by Id_X, such that for every object $\mathcal{Y} \in A^0$ one has

$$f \circ \mathrm{Id}_x = f, \quad \text{for every } f \in \mathrm{Mor}(\mathcal{X}, \mathcal{Y}),$$
$$\mathrm{Id}_x \circ g = g, \quad \text{for every } g \in \mathrm{Mor}(\mathcal{Y}, \mathcal{X}).$$

1.2. Example. A category the objects of which are any sets and morphisms — any mappings is called Ens.

1.3. Example. A category the objects of which are any topological spaces, and morphisms — continuous mappings is called Top.

1.4. Example. A category the objects of which are arbitrary, but finitely dimensional $\mathcal{C}^k$ differential manifolds and morphisms — arbitrary $\mathcal{C}^k$ diffeomorphisms is called $\mathcal{C}^k$Diff.[2]

1.5.–1.7. Examples. Analogously, one speaks about categories Vect, Gr, and Ab, the objects of which are vector spaces, groups, Abelian groups, and morphisms — any linear mappings and any homomorphisms, correspondingly.

1.8. Example. Let us recall that an *affine space* is a triple $\mathcal{A} = (\mathcal{E}, \mathcal{V}, +)$, where $\mathcal{E}$ is any set, $\mathcal{V}$ a vector space, and $+$ a mapping $+: \mathcal{E} \times \mathcal{V} \to \mathcal{E}$ such that $\mathcal{V}$ acts as a transitive and free group of transformations of $\mathcal{E}$ (for more details consult Schwartz, vol. I 1979, pp. 185–199). Let $\mathcal{A}_1 = (\mathcal{E}_1, \mathcal{V}_1, +)$ and $\mathcal{A}_2 = (\mathcal{E}_2, \mathcal{V}_2, +)$ be affine spaces. Let us define a mapping $f: \mathcal{E}_1 \to \mathcal{E}_2$ such that there exists a corresponding mapping $\tau f: \mathcal{V}_1 \to \mathcal{V}_2$ satisfying the condition

$$f(v + p) = \tau f(v) + f(p),$$

for any $p \in \mathcal{E}_1$ and $v \in \mathcal{V}_1$. The mapping f is said to be an *affine morphism*. The class of affine spaces together with affine morphisms as morphisms forms a category called Aff.

1.9. Example. A category, the objects of which are any metric spaces and morphisms — any contractions, i.e., mappings $\alpha: \mathcal{X} \to \mathcal{Y}$ such that

$$\rho_Y(\alpha(x_1), \alpha(x_2)) \leq \rho_X(x_1, x_2),$$

where ρ_X and ρ_Y are metrics on spaces $\mathcal{X}$ and $\mathcal{Y}$ correspondingly, $x_1, x_2 \in \mathcal{X}$, is called Metr.

1.10. Example. A category the objects of which are fibre bundles and morphisms — bundle morphisms (see chapter 3, sec. 2), is called Bun.

1.11. Definition. A category B is a *subcategory* of a category A if the objects and morphisms of B are also objects and morphisms of A, and if the compositions

[2] In the case of $\mathcal{C}^\infty$Diff, we shall simply write Diff.

of morphisms in B are restrictions of the corresponding compositions of morphisms in A.

For instance, Ab is a subcategory of Gr.

1.12. Definition. A morphism $f \in \mathrm{Mor}(\mathcal{X}, \mathcal{Y})$ is said to be an *isomorphism* if there exists another morphism $g \in \mathrm{Mor}(\mathcal{Y}, \mathcal{X})$ such that $f \circ g = \mathrm{Id}_Y$ and $g \circ f = \mathrm{Id}_X$.

For instance, bijections are isomorphisms in Top, isometries — in Metr, group isomorphisms — in Gr and Ab. Having a category A, one can define a category $\oint A$, the objects of which are those of A, and morphisms — the isomorphisms of A.

In the category theory, morphisms play, in a sense, a more important role than objects. One can construct a category axiomatically in such a way that its primitive concepts, besides logical constants and the equality sign, are (1) "α is a morphism", and (2) "composition of morphisms $\alpha \circ \beta$ is defined and equal γ". The axioms guarantee the associativity of compositions and the existence of identity morphisms. Since the concept of objects does not appear in the axioms, one speaks of an *objectless category theory*. It can be shown that a category defined at the beginning of the present section is a model of the axiomatic objectless category theory (see Semadeni and Wiweger 1978, pp. 44–46).

An important concept in the category theory is that of functor.

1.13a. Definition. Let A and B be two categories, A^0 and B^0 — the classes of their objects, and $\mathrm{Mor}(A)$ and $\mathrm{Mor}(B)$ — the classes of their morphisms. *A covariant functor* from the category A to the category B is a mapping

$$\mathcal{C}\colon (A, A^0, \mathrm{Mor}(A)) \longrightarrow (B, B^0, \mathrm{Mor}(B)),$$

satisfying the following conditions:

1) if $f \in \mathrm{Mor}(\mathcal{X}, \mathcal{Y}) \in \mathrm{Mor}(A)$, $\mathcal{X}, \mathcal{Y} \in A^0$, then $\mathcal{C}(f) \in \mathrm{Mor}(\mathcal{C}(\mathcal{X}), \mathcal{C}(\mathcal{Y})) \in \mathrm{Mor}(B)$, $\mathcal{C}(\mathcal{X}), \mathcal{C}(\mathcal{Y}) \in B^0$;
2) if $f, g \in \mathrm{Mor}(A)$ and f can be composed with g, then $\mathcal{C}(g) \circ \mathcal{C}(f) = \mathcal{C}(g \circ f) \in \mathrm{Mor}(B)$;
3) $\mathcal{C}(\mathrm{Id}_X) = \mathrm{Id}_{\mathcal{C}(X)}$, $\mathcal{X} \in A^0$, $\mathcal{C}(\mathcal{X}) \in B^0$, $\mathrm{Id}_X \in \mathrm{Mor}(A)$, $\mathrm{Id}_{\mathcal{C}(X)} \in \mathrm{Mor}(B)$.

The following diagram elucidates the structure of the above definition:

$$A\colon \mathcal{X} \xrightarrow{g} \mathcal{Y} \xrightarrow{f} \mathcal{Z};$$
$$B\colon \mathcal{C}(\mathcal{X}) \xrightarrow{\mathcal{C}(f)} \mathcal{C}(\mathcal{Y}) \xrightarrow{\mathcal{C}(g)} \mathcal{C}(Z)$$

The symbol $\mathcal{C}(f)$ is often replaced by the symbol f_*. To simplify the notation we shall also write $\mathcal{C}\colon A \to B$.

1.13b. Definition. *A contravariant functor* from the category A to the category B is obtained by introducing to definition 1.13a modifications visualised in the following diagram:

$$A\colon \mathcal{X} \xrightarrow{f} \mathcal{Y} \xrightarrow{g} \mathcal{Z};$$
$$B\colon \mathcal{C}(\mathcal{X}) \xleftarrow{\mathcal{C}(f)} \mathcal{C}(\mathcal{Y}) \xrightarrow{\mathcal{C}(g)} \mathcal{C}(\mathcal{Z})$$

The symbol $\mathcal{C}(f)$ is often replaced by the symbol f^*.

1.14. Example. A mapping

$$\mathcal{C}\colon (\mathcal{C}^k\mathrm{Diff}, (X^n, x), f) \longrightarrow (\mathcal{C}^{k-1}\mathrm{Diff}, \mathcal{T}_X(X), f'(x)),$$

where X^n is any n–dimensional $\mathcal{C}^k$ manifold, $\mathcal{T}_X(X^n)$ — the tangent space to X^n at a point $x \in X^n$, and f' — the derivative mapping of the mapping f transforming a $\mathcal{C}^k$ manifold into a $\mathcal{C}^k$ manifold, is a covariant functor from $\mathcal{C}^k\mathrm{Diff}$ to $\mathcal{C}^{k-1}\mathrm{Diff}$.

The category theory has enormous ability to order mathematical structures and make precise concepts which has so far been used intuitively. The notion of naturality belongs to such concepts. Eilenberg and MacLane considered the formalization of this notion as a main motive of creating the category theory (this can be seen in the title of their pioneering work; see Eilenberg and MacLane 1945). Even if further developments of the theory transcended this goal, it remains true for the present Appendix. Strict definition of the naturality concept is essential to grasp logical dependencies between different space–time structures.

1.15. Definition. Let $\mathcal{C}_1$ and $\mathcal{C}_2$ be two functors of the same variance (i.e., either both covariant or both contravariant) from a category A to a category B:

$$\mathcal{C}_1\colon A \longrightarrow B,$$
$$\mathcal{C}_2\colon A \longrightarrow B.$$

A *natural mapping* (called also a *morphism functor*) of $\mathcal{C}_1$ into $\mathcal{C}_2$ is a mapping

$$\mathcal{N}\colon A^0 \longrightarrow \mathrm{Mor}(B),$$

ascribing to every object $\mathcal{A} \in A^0$ a morphism of the category B:

$$\mathcal{A} \longmapsto \mathcal{N}(\mathcal{A}) \in \mathrm{Mor}(B),$$

in such a way that

$$\mathcal{N}(\mathcal{A})\colon \mathcal{C}_1(\mathcal{A}) \longrightarrow \mathcal{C}_2(\mathcal{A}).$$

It is also assumed that, for every $f\colon \mathcal{A}_1 \to \mathcal{A}_2$, $f \in \mathrm{Mor}(A)$, $\mathcal{A}_1$, $\mathcal{A}_2 \in A^0$, the following diagram is commutative (it is assumed that all the required compositions are meaningful)

$$\begin{array}{ccc} \mathcal{C}_1(\mathcal{A}_1) & \xrightarrow{\mathcal{C}_1(f)} & \mathcal{C}_1(\mathcal{A}_2) \\ \downarrow{\scriptstyle \mathcal{N}(\mathcal{A}_1)} & & \downarrow{\scriptstyle \mathcal{N}(\mathcal{A}_2)} \\ \mathcal{C}_2(\mathcal{A}_1) & \xrightarrow{\mathcal{C}_2(f)} & \mathcal{C}_2(\mathcal{A}_2) \end{array}$$

(all in this diagram occurs in the category B). If, additionally, for every $\mathcal{A} \in A^0$ the morphism $\mathcal{N}(\mathcal{A})$ is isomorphism in B, then the functors $\mathcal{C}_1$ and $\mathcal{C}_2$ are said to be *naturally equivalent*.

1.16. Example. Let us consider a contravariant functor $*$: Vect $\to$ Vect, defined by $f^*(g) = g \circ f$ for any mapping $f: \mathcal{A} \to \mathcal{B}$, $g: \mathcal{B} \to \mathbf{R}$ (f^* is called a *pull back* of the function g by f; see chapt. 1, sec. 4). Analogously, we define f^{**} (a pull back of a pull back), and a covariant functor **: Vect $\to$ Vect. Now, let $A = B =$ Vect, $\mathcal{C}_1 =$ Id, $\mathcal{C}_2 = {}^{**}$. The mapping $\mathcal{N}(\mathcal{V}): \mathcal{V} \to \mathcal{V}^{**}$, $\mathcal{V} \in \text{Vect}^0$ is a natural transformation as it can be easily seen from the commutative diagram, valid for any $f: \mathcal{V} \to \mathcal{U}$, $\mathcal{U} \in \text{Vect}^0$

$$\begin{array}{ccc} \mathcal{C}_1(\mathcal{V}) = \mathcal{V} & \xrightarrow{\mathcal{C}_1(f) = f} & \mathcal{C}_1(\mathcal{U}) = \mathcal{U} \\ \downarrow \mathcal{N}(\mathcal{V}) & & \downarrow \mathcal{N}(\mathcal{U}) \\ \mathcal{C}_2(\mathcal{V}) = \mathcal{V}^{**} & \xrightarrow{\mathcal{C}_2(f) = f^{**}} & \mathcal{C}_2(\mathcal{U}) = \mathcal{U}^{**} \end{array}$$

If we limit our considerations to the category of finitely–dimensional vector spaces, the functors $\mathcal{C}_1 =$ Id and $\mathcal{C}_2 = {}^{**}$ are naturally equivalent (see Trautman 1973b, p. 185).

Definitions of natural transformation and natural equivalence can be "naturally" extended to the case of functors of many variables (see Semadeni and Wiweger 1978 (pp. 134–136).

2. Aristotelian Space–Time

Dynamical theories investigate motions of bodies under the influence of forces. The first goal of a dynamical theory is to provide a criterion allowing one to determine whether any force has appeared. The role of such a criterion is usually played by the definition of "free motions", i.e., motions which are going on without any intervention of forces. A deviation from a free motion testifies to the appearance of a force. The formulation of such a criterion is the content of the *first law of dynamics*. If one assumes that motions under no influence of forces are determined by the structure of space–time — and this assumption seems to be very natural — then the first law of dynamics contains in itself information about this structure. Geometry, interpreted in such a way, becomes a part of the given dynamical theory, and, as it belongs to physics, is subject to empirical verification.

In Aristotle's dynamics[3] the first law can be formulated as follows:

There exists the natural rest,

or, equivalently,

[3]It is less known that Aristotle formulated a complete dynamics, and attempted at doing it in a quantitative manner (see Aristotle's *Physics*, especially the final parts of book VII). From our present point of view the Aristotelian "physics" is a mixture of philosophical views, purely verbal analyses, empirical statements (often erroneous ones), and a natural knowledge based on them. One could distillate from this a part which would correspond to what is today called physics. It goes without saying that the basic principles of the Aristotelian physics have been experimentally falsified.

a body, not acted upon by any force, remains in the absolute rest.[4]

This law implies the structure of space–time in the following way. The absolute rest allows one to meaningfully speak about the same place at different times (a body in the state of absolute rest **always** occupies the same place), and consequently it determines a projection π_Σ from the four–dimensional space–time $\mathcal{A}$ into a three–dimensional manifold Σ. The existence of the Aristotelian absolute time makes it possible to meaningfully speak of the absolute simultaneity of events (even if they are distant from each other), and consequently it defines a projection π_T of the manifold $\mathcal{A}$ into the "axis of absolute time" T. The existence of the projections π_Σ and π_T is equivalent to the existence of the absolute space and the absolute time, respectively. It follows that the Aristotelian space–time $\mathcal{A}$ is a product bundle ($\mathcal{A} = \Sigma \times T$, Σ, π_Σ).[5] One could easily show that this structure is natural for the Aristotelian space–time or, more strictly, that between suitably defined category of Aristotelian space–times and the category of product bundles there exists a natural transformation. Proof of this fact is similar to the proof which will be carried out in detail for space–time of classical mechanics.

Ascribing to the sets Σ, T and $\mathcal{A}$ the structure of smooth manifolds is a stylization of the history, but a stylization not too far exaggerated, if one takes into account the fact that the contemporary differential manifold notion directly comes from the Aristotelian concept of *continuum* for which ideas of space and time in turn constituted its prototypes.

The structure of the Aristotelian space–time is invariant with respect to the direct product of dilatation, translation and rotation groups (acting on Σ) by an affine group (acting on T).

The appearance of absolute velocity with respect to the absolute space testifies to the intervention of a force. In Aristotle's views, this force is proportional to the velocity produced by it. This statement can be thought of as to form the second law of the Aristotelian dynamics. However, we shall skip this problem as marginal for our discussion.[6]

3. Space–Time of Classical Mechanics Without Gravitational Field

Reconstruction of the space–time structure presupposed by classical mechanics is especially instructive. It not only shows that the structures, discovered later on by

[4]One should remember that Aristotle assumed the existence of the so–called natural motions (e.g. motions of heavy bodies towards the centre of the Earth). Of course, we could include such motions into our reconstruction of the Aristotelian dynamics, but this would highly complicate the picture without throwing any new light on questions we are interested in.

[5]One can also say that the Aristotelian space–time $\mathcal{A}$ is the Cartesian product $\mathcal{A} = \Sigma \times T$.

[6]More precise formulation of the second law of the Aristotelian dynamics would require a far–reaching historical stylization. It would mainly consist in translating some terms, used by Aristotle in their every–day meanings, into their counterparts elaborated much later in order to correctly express laws of dynamics. For instance, the Aristotelian term "efficient factor" should be replaced by "force", "heaviness" by "mass", etc. (See also Toulmin and Goodfield, 1961, pp. 93–100.)

relativity theory, existed and were well functioning from the very beginning of modern physics, but also demonstrates how quickly a physical theory gains independence from the philosophical views of its inventor and begins to lead an autonomous life on its own. Newton ascribed to space and time, which appeared in mechanics created by himself, Aristotelian properties, that is to say he treated them as absolute ones. It turns out that this remains correct as far as time is concerned, but is false with regard to space. One should distinguish the structure of space–time as postulated by Newton himself and the structure of space–time as presupposed by classical mechanics. The former is often called *Newton's space–time* (see Ehlers 1973), the latter *Galileo space–time* (see Kopczyński and Trautman 1981, Trautman 1970, 1973b). In the following we shall speak simply of *space–time of classical mechanics.* This seems to be justified since there is a necessity to distinguish the space–time structures of classical mechanics *without gravitational field and with gravitational field.* Now, we are going to analyse the first of these two cases, postponing the second one to the next section.

The first law of the Newtonian dynamics establishes a democracy among all inertial reference frame: none of such frames is privileged in any way. The standard of the absolute rest disappears from the theory. The Aristotelian projection π_Σ ceases to be a natural structure. Every observer, connected with a reference frame, has his private standard of rest ("with respect to his own inertial frame"); that is to say, his own projection corresponds to the Aristotelian projection π_Σ. However, absolute simultaneity and absolute acceleration remain valid in the classical world picture. Absolute simultaneity defines "spaces of equal time", and consequently leaves the Aristotelian projection π_Σ unchanged. Absolute acceleration connects the velocities at different points of space–time with one another: vanishing of acceleration allows one to meaningfully speak of "the same velocity at space–time points separated from one another by non–zero distances" (but not of the zero absolute velocity in different inertial frames). This leads to the so–called *teleparallelism*, i.e., to the concept of parallel vectors at points separate from one another in space–time.[7] Therefore, one must assume that space–time of classical mechanics (without gravitational field) is a four–dimensional differential manifold equipped with an affine connection which makes meaningful parallel transport of vectors (teleparallelism). It is natural to consider this connection in terms of a principal fibre bundle over space–time. Since any local inertial frame of reference can be identified with a basis at a given space–time point, the principal fibre bundle in question should be concretized to the bundle of frames over space–time. Let us consider this structure in more detail.

An affine space $\mathcal{A} = (\mathcal{E}, \mathcal{V}, +)$ (see example 1.8) equipped with a bilinear map-

[7]Let us turn our attention to the fact that the absolute character of acceleration in classical mechanics does not imply the existence of an absolute space. From the historical point of view, the appearance of "absolute accelerations" constituted the main Newton's argument on behalf of the existence of absolute space. This argument was the focus of heated discussions around the conceptual structure of classical mechanics (see Raine 1981a, p. 1157).

ping $h: \mathcal{V}^* \times \mathcal{V}^* \to \mathbf{R}^1$, such that h is a symmetric positively defined mapping of rank $n-1$, where n is the dimension of the vector space $\mathcal{V}$, will be called *space–time of classical mechanics without gravitational field* and denoted by $\mathcal{N}$. We shall write $\mathcal{N} = (\mathcal{E}, \mathcal{V}, +; h)$.

Let $\mathcal{N}_1 = (\mathcal{E}_1, \mathcal{V}_1, +; h_1)$ and $\mathcal{N}_2 = (\mathcal{E}_2, \mathcal{V}_2, +; h_2)$ be two such space–times. An affine morphism f will be called *Galileo's morphism* if $h_2 = h_1 \circ (\tau f)^*$, where the asterisk denotes the dual functor. The class of space–times of classical mechanics without gravitation with Galileo morphisms as morphisms is a category which will be denoted by Gal; it can be thought of as a subcategory of the category $\oint$Diff. A Galileo's morphism, which is an automorphism, is called a *Galileo's transformation.*

Let $\mathcal{S} \subset \mathcal{V}^*$ be defined as a subspace for which $h = 0$, and $\mathcal{S}\perp$ as a subspace of vectors orthogonal to $\mathcal{S}$. Let $\pi_T: \mathcal{E} \to \mathcal{E}/\mathcal{S} \equiv T$ be a canonical projection. Now, N can be presented as a fibre bundle $(\mathcal{E}, T, \pi_T, \mathcal{S})$, where $\mathcal{S}$ is a typical fibre. A functor $\sigma: \mathrm{Gal} \to \mathrm{Bun}$ can be defined from the category of space–times of classical mechanics without gravity to the category of bundles

$$\sigma(\mathcal{N}) = (\mathcal{E}, T, \pi_T, \mathcal{S}),$$
$$\sigma(f) = (f, a),$$

where $f: \mathcal{E}_1 \to \mathcal{E}_2$ is a Galileo's morphism and $a: T_1 \to T_2$ a uniquely determined mapping satisfying the condition: $\Pi_{T_1} \circ f = a \circ \pi_{T_2}$.

Now, let us define another functor $\omega: \mathrm{Gal} \to \mathrm{Bun}$ to every space–time $\mathcal{N} \in \mathrm{Gal}^0$ ascribing a product fibre $(T \times \mathcal{S}, T, \pi_T)$. It turns out that there is no natural transformation between functors σ and ω. This statement is mathematical counterpart of the intuitive statement that space–time of classical mechanics (without gravity) is not a Cartesian product of absolute time and absolute space.

The existence of teleparallelism in space–time determines horizontal cross–sections of the frame bundle over space–time. Indeed, all frames, resulting from a parallel transport of an arbitrarily chosen frame at a certain point of space–time, smoothly select a single point of space–time in each fibre of the frame bundle. Intuition suggests that in this way the natural product structure, lost in space–time, can be recovered in the fibre bundle of frames over this space–time. We will see that this is indeed the case.

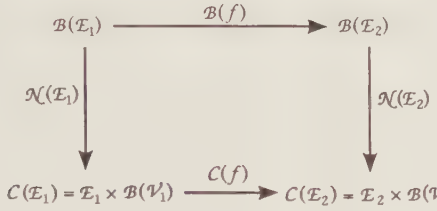
Let us define the following two functors: the functor

$$\mathcal{B}: \oint\mathrm{Aff} \longrightarrow P\mathrm{Bun}$$

from the category of affine spaces to the category of principal fibre bundles, associating with every object $\mathcal{A} = (\mathcal{E}, \mathcal{V}, +) \in \oint\mathrm{Aff}^0$ a frame bundle $\mathcal{B}(\mathcal{E}) \in P\mathrm{Bun}$, and the functor

$$\mathcal{C}: \oint\mathrm{Aff} \longrightarrow P\mathrm{Bun},$$

associating with every $\mathcal{A} = (\mathcal{E}, \mathcal{V}, +)$ a product bundle $\mathcal{C}(\mathcal{A}) = \mathcal{E} \times \mathcal{B}(\mathcal{V})$. These functors are naturally equivalent as it can be easily seen from the following commutative diagram for every $f: \mathcal{E}_1 \to \mathcal{E}_2$:



where $\mathcal{C}(f) = f \times \mathcal{B}(\tau f)$, and $\mathcal{N}$ is a natural transformation of the functor $\mathcal{B}$ into the functor $\mathcal{C}$. Therefore, in classical mechanics without gravity the product structure is naturally transferred from space-time to the frame bundle over space-time (see Trautman 1973b).

4. Space-Time of Classical Mechanics with Gravitational Field

At the basis of the Newtonian theory of gravity lies the statement which deserves the name *Galilean principle of equivalence*; it asserts that in a gravitational field all bodies fall with the same acceleration. As is well-known, this principle allows one to treat freely falling reference frames as *locally* non-accelerating, that is to say, reference frames in which laws of the Newtonian dynamics are locally valid (see Raine and Heller 1981, p. 83–84). However, the presence of a gravitational field causes focusing (or antifocusing) of space-time trajectories of freely falling bodies. For instance, the trajectories of bodies freely falling on Earth should meet at the point corresponding to the centre of the Earth. This effect is connected with the appearance of tidal forces. These forces liquidate the existence of teleparallelism: now, parallel transport depends on the path along which it is performed. We express this by saying that the affine connection over space-time of classical mechanics with gravity is non-integrable. This means that in the frame bundle over space-time there are no global cross-sections (in the case without gravity such cross-sections were implied by teleparallelism which now does not exist), and consequently the connection is not-flat (see chapt. 3, sec. 5). In other words: the presence of gravity manifests itself in the existence of a non-vanishing curvature form on the frame bundle over space-time.

The non-existence of global cross-sections in the frame bundle over space time suggests that this bundle, in the presence of a gravitational field, cannot be expressed in the form of a Cartesian product (globally). Indeed, in this case, there are no naturally equivalent functors associating with the space-time a trivial frame bundle. (See Raine 1981a, Raine and Heller 1981, chapter 6.)

5. The Special and General Theories of Relativity

Further logic of the evolution of space–time strucures is almost automatic. Space–time of special relativity (see chapter 2, sec. 3) has no product structure. Moreover, it has no natural fibration into global spaces of equal time (i.e., it has no projection π_T which would be independent of the frame of reference) that existed in space–time of classical mechanics.

Space–time of special relativity — in contradistinction to space–time of classical mechanics — carries a Lorentz structure and its affine connection is consistent with this structure. It is a flat connection which makes it possible to express the bundle of (orthonormal) frames in the form of a Cartesian product.

In general relativity (see chapter 4) the frame bundle over space–time has a non–vanishing curvature form which implies that this bundle is not, in any natural way, a trivial one. The essential difference between the space–time structure of classical mechanics with gravitational field and that of general relativity consists in that the latter is equipped with a Lorentz metric. In the architecture of the frame bundle this difference manifests itself in the existence of the soldering form which does not exist in classical mechanics.

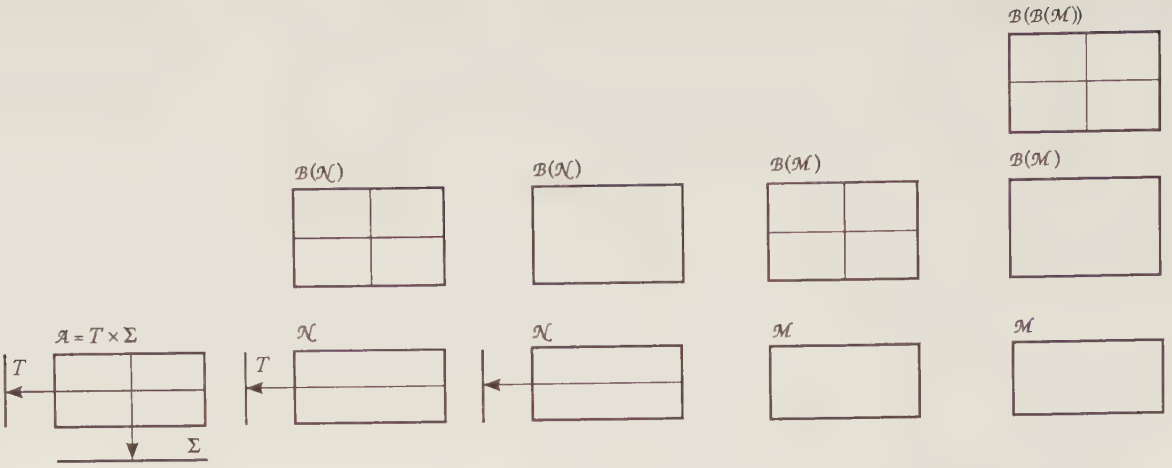
The existence of a linear connection in the frame bundle $\mathcal{B}(\mathcal{M})$ over space–time $\mathcal{M}$ is responsible for the fact that the frame bundle $\mathcal{B}(\mathcal{B}(\mathcal{M}))$ over the bundle $\mathcal{B}(\mathcal{M})$ is — in a natural way — a Cartesian product. This fact might suggest that the next step in the evolution of space–time theories could consist in depriving the bundle $\mathcal{B}(\mathcal{B}(\mathcal{M}))$ of its product structure and in the physical interpretation of this procedure (see Trautman 1970a, 1973b).

It is worth noticing that, when discussing the naturality problem in the case of general relativity, one should take into account the *category of principal fibre bundles with linear connections.* A *connection morphism* is a morphism of principal bundles such that the image under this morphism of any horizontal curve is a horizontal curve (see Sulanke and Wintgen 1977, p. 133).

The diagram on the adjacent page shows the logic of evolution of space–time structures (it is a modified version of the diagram appearing in the two works of Trautman quoted above).

6. Comments and Remarks

A. The logic of evolution of space–time theories, shown in the present Appendix, is a consequence of the strong stylization of the history of physics. It was the reconstruction of space–time structures of various theories in terms of modern differential geometry that has allowed us to see the logical connections between subsequent stages of the evolutionary process. The view that such a reconstruction reflects, at least some, real features of development of science seems well founded. On the other hand, however, there can be no doubt that the true history of science was much more complicated and much less logical: it was like a tree with the main stem and many



Space-time $\mathcal{A}$ of Aristotle has a product structure

Space-time $\mathcal{N}$ of classical mechanics without gravity has no product structure, but it admits fibration onto spaces of equal time

Frame bundle $\mathcal{B}(\mathcal{N})$ has a product structure

Space-time $\mathcal{N}$ of classical mechanics with gravity

Frame bundle $\mathcal{B}(\mathcal{N})$ has no product structure

$\mathcal{N}$ is an affine manifold

Space-time $\mathcal{M}$ of special relativity does not admit any natural fibration

Frame bundle $\mathcal{B}(\mathcal{M})$ has a product structure

Space-time $\mathcal{M}$ of general relativity

Frame bundle $\mathcal{B}(\mathcal{M})$ has no product structure

Frame bundle $\mathcal{B}(\mathcal{B}(\mathcal{M}))$ has a product structure

$\mathcal{M}$ is a Lorentz manifold

side–branches rather than a logical chain the parts of which smoothly change from one into another.

For example, the ideas of Berkeley and Leibniz, later on developed by Mach, constituted an important side–branch which could be thought of as a concurrence with respect to the Newtonian main stem (see Raine and Heller 1981, chapter 4; Barbour 1983). The essence of these ideas is contained in the postulate that the entire structure of space–time, and consequently the whole of dynamics, should be reduced to relationships between bodies without assuming any "absolute" manifold (see chapter 1, sec. 6, B). An attempt to construct such a dynamics has been undertaken by Barbour and Bertotti (1977). These authors believe that laws of the new dynamics should be invariant with respect to the following transformation group called the *Leibniz group*,

$$r \longmapsto \mathcal{A}(t)r + g(t),$$
$$t \longmapsto f(t),$$

where $\mathcal{A}(t)$ is an orthogonal matrix, and f and g are any smooth functions of time (additionally, one assumes that f is a monotonic function). The existence of the "universal time" t (the existence of the projection π_T) is justified by the Leibniz postulate of the "coexistence of events". The Leibniz group not only admits the rescaling of time, consisting in choosing new units and a new beginning of time counting (as did the Galileo group in classical mechanics), but also more general time transformations. The Leibniz group does not distinguish uniform motions; it also admits any suitably smooth rotations of instantaneous spaces. The Galileo group is a subgroup of the Leibniz group. (See also Liebscher and Yourgrau 1979).

B. In the philosophy of science there is a heated polemics in which the following views fight one another (a) the evolution of science goes on more or less logically, and the "time arrow" of this evolution is determined by the continuous accumulation of results (Popper, Lakatos); (b) strictly speaking, there is no evolution, but periods of "normal development" are separated from each other by all–changing revolutions; "normal periods" from before and after a revolution are "incommensurable" (Kuhn); (c) one should stop speaking about the progress in science since science is not ruled by logic but by laws of sociology and psychology of scientists (Wilson) or by the sole rule "anything goes" (Feuerabend). The polemics is often referred to in the literature as dispute on the rationality of science or of its evolution. The ideas briefly signaled in (c), which try to reduce all rationality criteria to social conditions changing from epoch to epoch, have recently grown in number and popularity (for a review of this dispute see for instance, Życiński 1988). The analyses, carried out in the present Appendix, throw some light on this problem. In spite of the strong historical stylization, which was indispensable in order to translate older space–time theories into the language of the present scientific paradigm, one cannot ignore the very fact that earlier theories **can be translated** into the modern language, and that in this process surprising logical dependencies are revealed within both the structures of

particular theories and the structure of the entire evolutionary chain. Of course, it is true that in this way our present criteria have been adapted to the former theories in such a manner that the whole procedure could succeed. Nevertheless, the result of this analysis should be considered as philosophically significant.

Bibliographical Note. Elements of the mathematical category theory could be learned, for instance, from Dodson (1980) or Semadeni and Wiweger (1972, 1978). Applications of this theory to differential geometry can be found in Sulanke and Wintgen (1972), and to the analysis of space–time structures in the works by Trautman (1970a,b, 1973b). Mathematical aspects of the classification of space–times (in terms of bundles which can be associated with the frame bundle over space–time) are presented in Trautman's paper (1976).

Reconstruction of space–time theories in the language of modern differential geometry has been carried out in the following works: Penrose 1968; Ehlers 1973; Trautman 1968, 1970a, 1971, 1973b; Kopczyński and Trautman 1981, Thirring 1978 (chapter 6); Raine and Heller 1981; Toretti 1983; Friedman 1983. Toretti's book contains some historical comments; Friedman focuses on philosophical aspects. One should take into account the different terminological conventions adopted by various authors, which sometimes could give the misleading impression that results of the analyses lead to contradictory results.

References

Abbot, L., 1988, The Mystery of the Cosmological Constant, *Scientific American* **258**, 82–88.

Abraham R., Marsden J., 1979, *Foundations of Mechanics*, Benjamin, New York (revised edition).

Adler, R., Bazin, M., Schiffer, M., 1965, *Introduction to General Relativity*, McGraw–Hill, New York, St. Louis ...

Aharoni, J., 1959, *The Special Theory of Relativity*, Clarendon Press, Oxford.

Arnold, W. I., 1974, *Mathematical Methods of Classical Mechanics*, Nauka, Moscow.

Auslander, L., Mac Kenzie, R. E., 1963, *Introduction to Differentiable Manifolds*, MacGraw–Hill, New York, San Francisco ...

Banach, S., 1932, *Theorie des opérations linéaires*, Warszawa.

Barbour, J. B., 1974, Relative–Distance Machian Theories, *Nature (Phys. Sci.)* **249**, 328–329 (corrected in *Nature* **250**, 1974, 606).

Barbour, J. B., 1983, Philosophical Principles and the Problem of Motion, *Philosophy in Science* **1**, 71–87.

Barbour, J., Bertotti, B., 1977, Gravity and Inertia in a Machian Framework, *Nuovo Cim.* **38B**, 1–27.

Barrow, J. D., 1978, Quiescent Cosmology, *Nature* **272**, 211–215.

Barrow, J. D., 1983, Anthropic Definitions, *Q. Jl. Royal Astron. Soc.* **24**, 146–153.

Barrow J. D., 1988, *The World Within the World*, Clarendon Press, Oxford.

Barrow, J. D., Matzner, R. A., 1977, The Homogeneity and Isotropy of the Universe, *Mont. Not. Royal Astron. Soc.* **181**, 719–727.

Barrow, J. D., Tipler, F. J., 1986, The Anthropic Cosmological Principle, Clarendon Press, Oxford.

Barrow, J. D., Turner, M. S., 1982, The Inflatory Universe — Birth, Death and Transfiguration, *Nature* **298**, 801–805.

Bass, R. W., Witten, L., 1957, Remark on Cosmological Models, *Rev. Mod. Phys.* **29**, 452–453.

Beem, J. K., Ehrlich, P. E., 1981, *Global Lorentzian Geometry*, Marcel Dekker, New York–Basel.

Bergamnn, P. G., 1976, *Introduction to the Theory of Relativity*, Dover Publ., New York.

Berstein, H. J., Philips, A. V., 1981, Fibre Bundles and Quantum Theory, *Scientific American* **245**, 94–109.

Berry, M., 1976, *Principles of Cosmology and Gravitation*, Cambridge University Press.

Bishop, R. L., Crittenden, R. J., 1964, *Geometry of Manifolds*, Academic Press, New York–London.

Bishop, R. L., Goldberg, S. I., 1980, *Tensor Analysis on Manifolds*, Dover Publ., New York.

Bishop, R. L., O'Neill. B., 1969, Manifolds of Negative Curvature, *Trans. Amer. Math. Soc.* **145**, 1–49.

Bondi, H., 1960, *Cosmology*, Cambridge University Press.

Bondi, H., Gold, T., 1948, The Steady–State Theory of the Expanding Universe, *Mont. Not. Royal Astron. Soc.* **108**, 252–270.

Bourbaki, N., 1969, *Elements d'histoire des mathematiques*, Hermann, Paris.

Brickell, F., Clark, R. S., 1970, *Differentiable Manifolds — An Introduction*, Van Nostrand, London, New York ...

Brill, D., 1982, Linearization Stability, in: *Space–time and Geometry*, ed.: R. A. Matzner, L. C. Shepley, University of Texas Press, Austin, 59–81.

Carmeli, M., 1977, *Group Theory and General Relativity*, McGraw–Hill, New York, London ...

Carr, J. B., 1982, The Anthropic Principle, *Acta Cosmol.* **11**, 143–151.

Carr, J. B., Rees, M. J., 1979, The Anthropic Principle and The Structure of the Physical World, *Nature* **278**, 605–612.

Cartan, E., 1922, Sur une generalisation de la notion de courbure de Riemann et les Espaces a Torsion, *Compt. Rend. Acad. Sci. (Paris)* **174**, 593–595.

Cartan, E., Sur les varietes a connexion affine et la theorie de la relativité genearalisée
1923, I: *Ann. Ec. Norm. Sup.* **40**, 325–412;
1924, I (suite): ibid. **41**, 1–25;
1925, II: ibid. **42**, 17–88.

Carter, B., 1974, Large Number Coincidences and the Anthropic Principle in Cosmology, in: *Confrontation of Cosmological Theories with Observational Data, IAU Symposium*, ed.: M. S. Longair, Reidel, Dordrecht–Boston, 291–298.

Choquet–Bruhat, Y., DeWitt–Morette,C., Dillard–Bleick, M., 1982, *Analysis, Manifolds and Physics* (revised edition), North–Holland, Amsterdam.

Clarke, C. J. S, 1970, On the Global Isometric Embedding of Pseudo–Riemannian Manifolds, *Proc. Royal Soc. London*, **A314**, 417–429.

Clarke, C. J. S., 1973, Local Extensions in Singular Space–Times, *Commun. math. Phys.* **32**, 205–214.

Clarke, C. J. S., 1975, The Classification of Singularities, *GRG* **6**, 35–40.

Clarke, C. J. S., 1976, Space–Time Singularities, *Commun. math. Phys. 49*, 17–23.

Clarke, C. J. S., Schmidt, B. G., 1977, Singularities: The State of the Art, *GRG* **8**, 129–137.

Collins, C. B., 1974, Tilting at Cosmological Singularities, *Commun. math. Phys.* **39**, 131–151.

Collins, C. B., 1977, The Role of Mathematics in Gravitational Physics — From the Sublime to the Subliminal? *GRG* **8**, 717–721.

Collins, C. B., Hawking, S. W., 1973, Why is the Universe Isotropic? *Astrophys. J.* **180**, 317–334.

Daniel, M., Viallet, C. M., 1980, The Geometrical Setting of Gauge Theories of the Yang–Mills Type, *Rev. Mod. Phys.* **52**, 175–197.

Demaret, J., Barbier, C., 1981, Le principe anthropique en cosmologie, *Rev. Quest. Sci.* **151**, 181–222; 461–509.

Demiański, M., 1985, *Relativistic Astrophysics*, Polish Scientific Publishers — Pergamon Press, Warszawa, Oxford ...

Dicke, R. H., 1961, Dirac's Cosmology and Mach's Principle, *Nature* **192**, 440–441.

Dicke, R. H., Peebles, P. J. E., Roll, P. G., Wilkinson, D. T., 1965, Cosmic–Black–Body Radiation, *Astrophys. J.* **142**, 114–119.

Dixon, W. G., 1978, *Special Relativity — The Foundation of Macroscopic Physics*, Cambridge University Press.

Dodson, C. T. J., 1980, *Categories, Bundles and Spacetime Topology*, Shiva Publishing, Orpington.

Dolgov, A. D., Zeldovich, I. B., Sagin, M. B., 1988, *Cosmology of the Early Universe*, Moscow University Press (in Russian).

Drechsler, W., Mayer, M. E., 1977, *Fiber Bundle Techniques in Gauge Theories*, Springer, Berlin — Heidelberg — New York.

Eddington, A. S., 1923, *The Mathematical Theory of Relativity*, Cambridge University Press.

Eguchi, T., Gikey, P. B., Hanson, A. J., 1980, Gravitation, Gauge Theories and Differential Geometry, *Phys. Rep.* **66**, 215–393.

Ehlers, J., 1973, The Nature and Structure of Space–Time, in: *The Physicist's Conception of Nature*, ed.: J. Mehra, Reidel, Dordrecht, 71–91.

Ehlers, J., Pirani, F. A. E., Schild, A., 1972, The Geometry of Free Fall and Light Propagation, in: *General Relativity — Papers in Honour of J. L. Synge*, ed.: L. O'Raifeartaigh, Clarendon Press, Oxford, 65–83.

Eilenberg, S., MacLane, S., 1945, General Theory of Natural Equivalencies, *Trans. Amer. Math. Soc.* **58**, 231–294.

Einstein, A., 1905, Zur Elektrodynamik bewegter Körper, *Ann. Phys.* **17**, 891–921; English translation: *The Principle of Relativity*, Dover Publ., 1923, 35–65 and in: Miller 1981, 391–415.

Einstein, A., 1917, Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie, *Sitzungsber. preuss. Akad. Wiss.* **1**, 142–152; English translation: *The Principle of Relativity*, Dover Publ., 1923, 177–188.

Ellis, G. F. R., 1971, Topology and Cosmology *GRG*, **2**, 7–21.

Ellis, G. F. R., 1975, Cosmology and Verifiability, *Q. Jl. R. Astr. Soc.* **16**, 245–264.

Ellis, G. F. R., 1980, Limits to Verification in Cosmology, *Ann. New York Acad. Sci.* **336**, 130–160.

Ellis, G. F. R., 1984, Relativistic Cosmology: Its Nature, Aims and Problems, in: *General Relativity and Gravitation, Invited Papers and Discussion Reports*, Padua, July 3–8, 1983, ed.: B. Bertotti, F. de Felice, A. Pascolini, Reidel, Dordrecht — Boston, 215–288.

Ellis, G. F. R., Maartens, R., Nel, S. D., 1978, The Expansion of the Universe, *Mont. Not. Royal Astron. Soc.* **184**, 439–465.

Ellis, G. F. R., Maartens, R., Nel, S. D., Stoeger W. R., Whitman A. P., Ideal Observational Cosmology, *Physic Reports* **124**, 315–417.

Ellis, G. F. R., Perry, J. J., 1979, Towards a "Correctionless" Observational Cosmology, *Mont. Not. Royal Astron. Soc.* **187**, 357–370.

Ellis, G. F. R., Schmidt, B. G., 1977, Singular Space–Times, *GRG* **11**, 915–953.

Ellis, G. F. R., Sciama, D. W., 1972, Global and Non–Global Problems in Cosmology, in: *General Relativity — Papers in Honour of J. L. Synge*, ed.: L. O'Raifeartaigh, Clarendon Press, Oxford, 35–59.

Ellis, G. F. R., Stoeger, W., 1988, Horizons in Inflationary Universes, *Class. Quantum Grav.* **5**, 207–220.

Engelking, R., 1975, *General Topology*, Polish Scientific Publishers, Warsaw (in Polish).

Fisher, A., Marsden, J. E., 1979, The Initial Value Problem and the Dynamical Formulation of General Relativity, in: *General Relativity — An Einstein Centenary Survey*, eds.: S. W. Hawking, W. Israel, Cambridge University Press, 138–211; 850–859.

Fisher, A., Marsden, J. E., Moncrief, V., 1980, Symmetry Breaking in General Relativity, in: *Essays in General Relativity — A Festschrift for Abraham Taub*, ed.: F. J. Tipler, Academic Press, New York, London ..., 79–96.

Flanders, H., 1963, *Differential Forms with Applications to the Physical Sciences*, Academic Press, New York — London.

Friedman, M., 1983, *Foundations of Space–Time Theories — Relativistic Physics and Philosophy of Science*, Princeton University Press.

Friedrich, H., Steward, J. M., 1982, *Characteristic Initial*
Data and Wave Front Singularities in General Relativity, Preprint, Max–Planck–Institut, München.

Geroch, R., 1970, Singularities, in: *Relativity* ed.: S. Fickler, M. Carmeli, L. Witten, Plenum, New York, 259–291.

Geroch, R., 1971, Space–Time Structure from a Global Viewpoint, in: *General Relativity and Cosmology, Enrico Fermi XLVII Course*, Academic Press, Chicago — London.

Geroch, R., 1978, *General Relativity from A to B*, The University of Chicago Press, Chicago — London.

Geroch, R., Horowitz, G. T., 1979, Global Structure of Space–Time, in: *General Relativity — An Einstein Centenary Survey*, ed.: S. W. Hawking, W. Israel, Cambridge University Press, 212–293; 859–860.

Godart, O., Heller, M., 1985, *Cosmology of Lemaître*, Pachart, Tucson.

Goenner, H., 1970, Mach's Principle and Einstein's Theory of Gravitation, in: *Boston Studies in the Philosophy of Science*, vol. VI: Ernst Mach — Physicist and Philosopher, Reidel, Dordrecht, 200–215.

Golda, Z., Heller, M., Szydlowski, M., 1983, Structurally Stable Approximations to Friedmann–Lemaître World Models, *Astrophys. Space Sc.* **90**, 313–326.

Groth, E. J., Peebles, J. E., Seldner, M., Soneira, R. M., 1977, The Clustering of Galaxies, *Scientific American* **237**, 5, 76–98.

Gruszczak, J., Heller, M., Multarzyński, P., 1988, A Generalization of Manifolds as Space–Time Models, *J. Math. Phys.* **29**, 2576–2580.

Gruszak, J., Heller, M., Multarzyński, P., 1989, Physics With and Without the Equivalence Principle, *Foundations of Physics* **19**, 607–618.

Guth, A. H., 1981, Inflationary Universe: A Possible Solution to the Horizon and Flatness Problems, *Phys. Rev.* **23D**, 347–356.

Harrison, E. R., 1981, *Cosmology — The Science of the Universe*, Cambridge University Press.

Hartle, J. B., Hawking, S. W., 1983, Wave Function of the Universe, *Phys. Rev.* **D28**, 2960–2975.

Hawking, S. W., 1971, Stable and Generic Properties in General Relativity, *GRG* **1**, 393–400.

Hawking, S. W., 1979, The Path–Integral Approach to Quantum Gravity, in: *General Relativity — An Einstein Centenary Survey*, eds.: S. W. Hawking, W. Israel, Cambridge University Press, 746–789; 894–895.

Hawking, S. W., 1982, The Boundary Conditions of the Universe, in: *Astrophysical Cosmology*, eds.: H. A. Bruck, G. V. Coyne, M. S. Longair, Pontificiae Academiae Scientiarum Scripta Varia, vol. 48, 563–574.

Hawking, S. W., 1984a, The Quantum State of the Universe, *Nuclear Physics* **B239**, 257–276.

Hawking, S. W., 1984b, Quantum Cosmology, in: *Relativity,*

Groups and Topology II, eds.: B. S. DeWitt, R. Stora, North–Holland, Amsterdam, Oxford ..., 332–379.

Hawking, S. W., Ellis, G. F. R., 1973, *The Large Scale Structure of Space–Time*, Cambridge University Press.

Hawking, S. W., Moss, I. G., 1982, Supercooled Phase Transitions in the Very Early Universe, *Phys. Lett.* **110B**, 35–38.

Heckmann, O., Schücking, I., 1959, Newtonsche und Einsteinsche Kosmologie, in: *Handbuch der Physik*, Band LII, Astrophysik IV: Sternsysteme, Springer, Berlin, Gottingen ..., 489–519.

Hehl, F. W., Spin and Torsion in General Relativity

1973, I: Foundations, *GRG* **4**, 333–349, 1974

1974, II: Geometry and Field Equations, *GRG* **5**, 491–516.

Heidmann, J., 1973, *Introduction à la Cosmologie*, Press Universitaires de France.

Heller, M., 1975a, The Influence of Mach's Thought on Contemporary Relativistic Physics, *Organon* **11**, 271–283.

Heller, M., 1977, Space–Time Structures, *Acta Cosmol.* **6**, 109–128.

Heller, M., 1978, Local — Large Scale — Global: On Certain

Methodological Questions of Cosmology, *Acta Cosmol.* **7**, 83–99.

Heller, M., 1981a, The Manifold Model for Space–Time, *Acta Cosmol.* **10**, 33–51.

Heller, M., 1981b, Relativistic Model for Space–Time, *Acta Cosmol.* **10**, 53–69.

Heller, M., 1981c, The Origin of Time, in: *The Study of Time IV*, ed.: J. T. Fraser, N. Lawrence, D. Park, Springer, New York, Heidelberg ..., 90–93.

Heller, M., 1982, Matter and Radiation Models in the Standard Universe, *Acta Cosmol.* **11**, 11–25.

Heller, M., 1988, Between Newton and Einstein, in: *Newton and the New Direction in Science, Proc. of the Cracow Conference, 25–28 May 1987*, eds.: G. V. Coyne, M. Heller, J. Życinski, Specola Vaticana, 155–173.

Heller, M., Multarzyński, P., Sasin, W., 1989, Algebraic Approach to Space–Time Geometry, *Acta Cosmol.* **16**, 53–85.

Hicks, N. J., 1965, *Notes on Differential Geometry*, Van Nostrand, Princeton.

Higgs, P. W., 1964, Broken Symmetries and The Masses of Gauge Bosons, *Phys. Rev. Lett.* **13**, 508–509.

Higgs, P. W., 1966, Spontaneous Symmetry Breakdown without Massless Bosons, *Phys. Rev.* **145**, 1156–1163.

Hoyle, F., 1948, A New Model for the Expanding Universe, *Mont. Not. Royal Astron. Soc.* **108**, 372–382.

Hubble, E., 1929, A Relation between Distance and Radial Velocity among Extra–Galactic Nebulae, *Proc. Nat. Acad. Sc.* **15**, 168–173.

Hubble, E., 1936, *The Realm of the Nebulae*, Yale University Press.

Hubble, E., 1937, *The Observational Approach to Cosmology*, Clarendon Press, Oxford.

Husemoller, D., 1975, *Fibre Bundles*, 2 edition, Springer, New York.

Kazanas, D., 1980, Dynamics of the Universe and Spontaneous Symmetry Breaking, *Astrophys. J.* **241**, L59–L63.

Kibble, T. W. B., 1961, Lorentz Invariance in Gravitational Field, *J. Math. Phys.* **2**, 212–221.

Kobayashi, S., Nomizu, K., 1963, *Foundations of Differential Geometry*, Interscience, New York — London.

Kramer, D., Stephani, H., Herlt, E., MacCallum, M., 1980, *Exact Solutions of Einstein's Field Equations*, ed.: E. Schmutzer, Cambridge University Press.

Kristian, J., Sachs, R. K., 1966, Observations in Cosmology, *Astrophys. J.* **143**, 379–399.

Landau, L., Lifszic, E., 1973, *The Field Theory*, Nauka, Moscow, (in Russian).

Landsberg, P. T., 1977, *Mathematical Cosmology — An Introduction*, Clarendon Press, Oxford.

Lemaître, G., 1927, Un univers homogène de masse constante et de rayon croissant, renadnt compte de la vitesse radiale des nébuleuses extra–galactiques, *Ann. Soc. Sc. Bruxelles* **47A**, 49–59; English translation: *Mont. Not. Royal Astron. Soc.* **91**, 1931, 483–490.

Lemaître, G., 1950, *The Primeval Atom — An Essay on Cosmology*, Van Nostrand, Toronto — New York — London.

Lemaître, G., 1972, *L'hypothese de l'Atome Primitif*, Éd. Culture et Civilisation, Bruxelles.

Liebscher, D–E., Yourgrau, W., 1979, Classical Spontaneous Breakdown of Symmetry and Induction of Inertia, *Annalen der Phys. (Leipzig)*, **36**, 20–24.

Lightman, A. P., Press, W. H., Price, R. H., Teukolsky, S. A., (1975), *Problem Book in Relativity and Gravitation*, Princeton Univ. Press.

Linde, A. D., A New Inflationary Universe Scenario: a Possible Solution of the Horizon, Flatness, Homogeneity, Isotropy, and Primordial Monopole Problems, *Phys. Lett.* **108B**, 398–395.

Longair, M., (ed.), 1974, *Confrontation of Cosmological Theories with Observational Data (IAU Symposium No 63)*, Reidel, Dordrecht — Boston.

Lubkin, E., 1963, Geometric Definition of Gauge Invariance, *Ann. Phys. (New York)* **23**, 233–283.

MacCallum, M. A. H., 1973, Cosmological Models from a Geometric Point of View, in: *Cargése Lectures*, vol. 6., ed.: E. Schatzman, Gordon and Breach, New York, 61–174.

Marsden, J. E.,1981, *Lectures on Geometric Methods in Mathematical Physics*, Society for Industrial and Applied Mathematics, Philadelphia.

Maurin, K., 1976, *Analysis*, Polish Scientific Publishers — Reidel, Warszawa — Dordrecht:

1976, part 1: *Elements*,

1980, part 2: *Integration, Distributions, Harmonic Functions, Tensor and Harmonic Analysis*.

Marleau–Ponty, J., 1965, *Cosmologie du XX^e siècle*, Éd. Gallimard.

Miller, A. I., 1981, *Albert Einstein's Special Theory of Relativity — Emergence (1905) and Early Interpretation (1905–1911)*, Addison–Wesley, London, Amsterdam ...

Milne, E. A., 1935, *Relativity, Gravitation and World Structure*, Clarendon Press, Oxford.

Minkowski, H., 1907, Das Relativitätsprinzip, a paper delivered on 5 of November 1907 during the meeting of the Math. Gesel. Göttingen, published in: *Ann. Phys.*, **47**, 1915, 927–938.

Minkowski, H., 1908a, Die Grundgleichungen für die elektromagnetischen Vorgänge in bewegten Körpern, *Göttinger. Nachr.* 53–111.

Minkowski, H., 1908b, Raum und Zeit, a paper delivered on 21 of September 1908 in Cologne, published in: *Phys. Z.*, **20**, 1909, 104–111; English Translation: *The Principle of Relativity*, Dover Publ., 73–96.

Misner, C. W., 1969, Mix–Master Universe, *Phys. Rev. Lett.* **22**, 1071–1074.

Misner C. W., 1980, Symmetry Paradoxes and other Cosmological Comments, in: *Some Strangeness in the Proportion: A Centennial Symposium to Celebrate the Achievement of Albert Einstein*, Addison–Wesley, 405–415.

Misner, C. W., Thorne, K., Wheeler, J. A., 1973, *Gravitation*, Freeman, San Francisco.

Moncrief, V., Space–Time Symmetries and Linearization Stability of the Einstein Equations, *J. Math. Phys.*:

1975; I: **16**, 493–498,

1976, II: **17**, 1893–1902.

Nash, C., Sen, S., 1983, *Topology and Geometry for Physicists*, Academic Press, London, Orlando ...

O'Neill, B., 1966, *Elementary Differential Geometry*, Academic Press, New York — London.

O'Neill, B., 1983, *Semi–Riemannian Geometry with Applications to Relativity*, Academic Press, New York — London.

North, J. D., 1965, *The Measure of the Universe*, Clarendon Press, Oxford.

Pacholczyk, A. G., 1984, *The Catastrophic Universe*, Pachart, Tucson.

Peebles, P. J. E., 1971, *Physical Cosmology*, Princeton University Press.

Peebles, P. J. E., 1980, *The Large–Scale Structure of the Universe*, Princeton University Press.

Penrose, R., 1968, Structure of Space–Time, in: *Battelle Rencontres* 1967, eds.: C. M. DeWitt, J. A. Wheeler, Benjamin 121–235.

Penrose, R., 1980, Null Hypersurface Initial Data for Classical Fields of Arbitrary Spin and for General Relativity, *GRG* **12**, 225–264.

Penrose, R., Rindler, W., 1984, *Spinors and Space–Time*, vol. 1: *Two–Spinor Calculus and Relativistic Fields*, Cambridge University Press.

Penzias, A. A., Wilson, R. W., 1965, A Measurement of Excess Antenna Temperature at 4080 Mc/s., *Astrophys. J.* **142**, 419–421.

Petrosian, V., 1974, Confrontation of Lemaitre Models and the Cosmological Constant with Observations, in: *Confrontation of Cosmological Theories with Observational Data (IAU Symposium No 63)*, Ed.: M. Longair, Reidel, Dordrecht — Boston, 31–46.

Pietrov, A. Z., 1966, *New Methods in General Relativity*, Nauka, Moscow (in Russian).

Raine, D. J., 1975, Mach's Principle in General Relativity, *Mont. Not. Royal Astron. Soc.* **171**, 507–528.

Raine, D. J., 1981a, Mach's Principle and Space–Time Structure, *Rep. Prog. Phys.* **44**, 1152–1195.

Raine, D. J., 1981b, *The Isotropic Universe*, Adam Hilger, Bristol.

Raine, D. J., Heller, M., 1981, *The Science of Space–Time*, Pachart, Tucson.

Reinhardt, M., 1973, Mach's Principle — A Critical Review, *Zeitschr. für Naturforschung*, **28a**, 529–537.

Richtmyer, R. D., *Principles of Advanced Mathematical Physics*, Springer, New York — Berlin:

1978: vol. 1,

1981: vol. 2.

Riemann, G. F. B., 1868, Über die Hypothesen, welche der Geometrie zu Grunde liegen (Habilitationsschrift 1854). *Abhandl. kgl. Ges. Wiss. zu Göttingen* **13**.

Riesz, F., Sz–Nagy, B., 1972, *Leçons d'analyse fonctionelle*, Akadémiai Kiadó, Budapest.

Rindler, W., 1956, Visual Horizons in World Models, *Mont, Not. Royal Astron. Soc.* **116**, 662–677.

Rindler, W., 1969, *Essential Relativity: Special, General, and Cosmological*, Van Nostrand, New York, Cincinnati ...

Rindler, W., 1977, *Essential Relativity — Special, General, and Cosmological, Springer*, New York, Heidelberg, Berlin.

Robertson, H. P., 1933, Relativistic Cosmology, *Rev. Mod. Phys.* **5**, 62–90.

Robertson, H. P., Noonan, T. W., 1968, *Relativity and Cosmology*, Saunders, Philadelphia, London ...

Roman, P., 1975, *Some Modern Mathematics for Physicists and Other Outsiders*, vol. 1 and 2, Pergamon, New York, Toronto ...

Rowan–Robinson, M., 1977, *Cosmology*, Clarendon Press, Oxford.

Ross, G. G., 1981, Unified Fields Theories, *Rep. Prog. Phys.* **44**, 655–718.

Ryan, M. P., Shepley, L. C., 1975, *Homogeneous Relativistic Cosmologies*, Princeton University Press.

Sachs, R. K., Wu, H., 1977a, General Relativity and Cosmology, *Bull. Amer. Math. Soc.* **83**, no. 6, 1101–1164.

Sachs R. K., Wu, H., 1977b, *General Relativity for Mathematicians*, Springer, New York.

Salam, A., 1968, in: *Elementary Particle Theory*, ed.:
N. Svartholm, Almquist and Wiksell Förlag, Stockholm.

Sandage, A. R., 1961, The Ability of the 200–inch Telescope to Discriminate between Selected World Models, *Astrophys. J.* **133**, 335–392.

Schmidt, B., 1972, Differential Geometry from a Modern Standpoint, in: *Relativity, Astrophysics and Cosmology*, ed.: W. Israel, Reidel, Dordrecht — Boston, 289–322.

Schutz, B. F., 1984, *Geometrical Methods of Mathematical Physics*, Cambridge University Press
1967, *Analyse mathématique*, Hermann, Paris, vol. 1 and 2.

Schutz, B. F., 1985, *A First Course in General Relativity*, Cambridge University Press.

Schwinger, J., The Theory of Quantized Fields, *Phys. Rev.*:
I: 1951, **82**, 914–927,
II: 1953, **91**, 713–728,
III: 1953, **91**, 728–740.

Sciama, D. W., 1962, On the Analogy between Charge and Spin in General Relativity, in: *Recent Developments in General Relativity*, Pergamon Press — Polish Scientific Publishers, 415–439.

Sciama, D. W., 1971, *Modern Cosmology*, Cambridge University Press.

Shapiro, I. I., 1979, *Experimental Tests of the General Theory of Relativity and Gravitation*, vol. 2, ed.: A. Held, Plenum Press, New York — London, 469–489.

Silk, J., Barrow, J. D., 1983, *The Left Hand of Creation*, Basic Books.

Silk, J., Szalay, A. S., Zeldovich, Y. B., 1983, The Large–Scale Structure of the Universe, *Scientific American*, **249**, 56–64.

Silk, J., 1980, *The Big Bang — The Creation and Evolution of the Universe*, Freeman and Comp.

Smale, S., 1980, *The Mathematics of Time*, Springer, New York — Heidelberg — Berlin.

Spanier, E. H., 1966, *Algebraic Topology*, McGraw–Hill, New York — San Francisco.

Steenrod, N., 1951, *The Topology of Fibre Bundles*, Princeton University Press.

Sulanke, R., Wintgen, P., 1972, *Differential Geometrie und Faserbündel*, VEB Deutscher Verlag der Wissenschaften, Berlin.

Synge, J. L., 1965, *Relativity: The Special Theory*, North–Holland, Amsterdam.

Szydłowski, M., Heller, M., Golda, Z., 1984, Structural Stability Properties of Friedman Cosmology, *GRG* **16**, 877–890.

Taylor, E. P., Wheeler, J. A., 1966, *Spacetime Physics*, Freeman and Comp., San Francisco — London.

Taylor, E. P., 1978, *Gauge Theories of Weak Interactions*, Cambridge University Press.

Thirring, W., 1978, *A Course in Mathematical Physics*, vol. 1: *Classical Dynamical Systems*, Springer, New York — Wien.

Thom, R.,1977, *Stabilité structurelle et morphogénèse*, Interéditions.

Tolman, R. C., 1934, *Relativity, Thermodynamics and Cosmology*, Clarendon Press, Oxford.

Tonnelat, M. A., 1971, *Histoire du principle de relativité*, Flammarion, Paris.

Toretti, R., 1983, *Relativity and Geometry*, Pergamon Press, Oxford, New York ...

Trautman, A., 1964, Foundations and Current Problems of General Relativity, in: *Lectures on General Relativity*, Brandeis Summer Institute in Theoretical Physics, Prentice–Hall, Englewood Cliffs, 1–248.

Trautman, A., 1970a, Fibre Bundles Associated with Space–Time, *Rep. Math. Phys.* **1**, 29–62.

Trautman, A., 1970b, Riemannian Bundles, *Bull. Acad. Pol. Sc., Sér. sc. math., astr. et phys.* **18**, 667–672.

Trautman, A., On the Einstein–Cartan Equations, *Bull. Acad. Pol. Sc., Sér, sc. math., astr. et phys.*:

I: 1972a, **20**, 185–190,

II: 1972b, **20**, 503–506,

III: 1972c, **20**, 859–896,

IV: 1973a, **21**, 345–346.

Trautman, A., 1973b, Theory of Gravitation, in: *The Physicist's Conception of Nature*, ed.: J. Mehra, Reidel, 179–198.

Trautman, A., 1973c, On the Structure of Einstein–Cartan Equations, *Symposia Mathematica* **12**, 139–162.

Trautman, A., 1975, Recent Advances in Einstein–Cartan Theory of Gravity, *Ann. N. Y. Acad. Sc.* **262**, 241–245.

Trautman, A., 1976, A Classification of Space–Time Structures, *Rep. Math. Phys.* **10**, 297–310.

Trautman, A., 1979, The Geometry of Gauge Fields, *Czech. J. Phys.* **B29**, 107–116.

Trautman, A., 1980, Fiber Bundles, Gauge Fields, and Gravitation, in: *General Relativity and Gravitation*, vol. 1, ed.: A. Held, Plenum Press, New York — London, 287–308.

Trautman, A., 1981, Geometrical Aspects of Gauge Configurations, *Acta Phys. Austriaca*, Suppl. XXIII, 401–432.

Trautman, A., 1984, *Differential Geometry for Physicists — Stony Brook Lectures*, Bibliopolis, Napoli.

Trefil, J. S., 1983, *The Moment of Creation*, Scribner's Sons, New York.

Tropp, E. A., Frenkel, W. Ya., Chernin, A. D., 1988, *Alexander Alexandrovitch Friedman — Life and Work*, Nauka, Moscow (in Russian).

[Understanding ...], 1979, *Understanding Space and Time*, The Open University Press.

Utiyama, R., 1956, Invariant Theoretical Interpretation of Interaction, *Phys. Rev.* **101**, 1597–1607.

Wallace, A. H., 1968, *Differential Geometry – First Steps*, W. A. Benjamin, New York — Amsterdam.

Weinberg, S., 1967, A Theory of Leptons, *Phys. Rev. Lett.* **19**, 1264–1266.

Weinberg, S., 1972, *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity*, E Wiley and Sons, New York, London ...

Weinberg, S., 1977, *The First Three Minutes*, A. Deutsch, London.

Weizsäcker von C. F., 1972, *Die Einheit der Natur*, Carl Hauser Verlag.

Weyl, H., 1918, Gravitation und Elektrizität, *Sitzungsber. Preuss. Acad. Wiss.* 465–480.

Weyl, H., 1919, Eine neue Erweterung der Relativitätstheorie, *Ann. Phys. Leipzig* **59**, 101–133.

Weyl, H., 1922, *Space–Time–Matter*, Dover Publ.

Weyl, H., 1955, *The Concept of a Riemann Surface*, Addison Wesley, Reading Mass. German original published in 1913.

Whitney, H., 1936, Differentiable Manifolds, *Ann. Math.* **37**, 645–680.

Will, C. M., 1979, The Confrontation between Gravitation Theory and Experiment, in: *General relativity — An Einstein Centenary Survey*, ed.: S. W., Hawking, W., Israel, Cambridge University Press, 24, 89, 833–846.

Will, C. M., 1979, The Confrontation between Gravitation Theory and Experiment, in: *General relativity — An Einstein Centenary Survey*, ed.: S. W., Hawking, W., Israel, Cambridge University Press, 24–89, 833–846.

Wolf, J. A., 1967, *Spaces of Constant Curvature*, McGraw–Hill, New York, St. Louis ...

Yang, C. N., Mills, R. L., 1954, Conservation of Isotopic Spin and Isotopic Gauge Invariance, *Phys. Rev.* **96**, 191–195.

Yasskin, Ph. B., 1979, *Metric–Connection Theories of Gravity*, University of Maryland, Dept. of Phys. and Astron., Thesis.

Zeldovitch, Ya, B., 1975, *The Structure and Evolution of the Universe*, Nauka, Moscow (in Russian).

Życinski, J., 1988, *The Structure of the Metascientific Revolution*, Pachart, Tucson.

Symbol Index

Chapter 1

(U, ϕ), (U_i, ϕ_i)	local charts, 3, 5
$\{A\}$	atlas
$\{\mathcal{M}\} = (X, \{A\})$	differential manifold, 6–7
$\psi \circ f \circ \phi^{-1}$	mapping f in charts (U, ϕ) and (W, ψ), 8
$\mathcal{G}(A)$	graph of a manifold $\mathcal{M} = (X, \{A\})$, 10
$\mathcal{V}$	vector space, 13
$\mathcal{V}^*$	dual space to $\mathcal{V}$, 13
$\mathbb{L}(\{\mathcal{V}, \mathcal{W}\})$	set of all linear mappings from $\mathcal{V}$ into $\mathcal{W}$, 13
$\mathcal{T}_x(\mathcal{M})$	tangent space to $\mathcal{M}$ at x, 11
$\mathcal{T}_x^*(\mathcal{M})$	dual space (cotangent space) to $\mathcal{T}_x(\mathcal{M})$, 13
f'	derivative of a mapping f, 13
f^*	pull back of a function f, 13
$f^*(\omega)$	pull back of a p–form ω, 14
$\tau_1 \otimes \cdots \otimes \tau_n$	tensor product, 13
$f\colon \underbrace{\mathcal{V}^* \times \cdots \times \mathcal{V}^*}_{r \text{ times}} \times \underbrace{\mathcal{V} \times \cdots \times \mathcal{V}}_{s \text{ times}} \to \mathbf{R}^1$	tensor of the type (r, s), 13
$\underbrace{\mathcal{V} \otimes \cdots \otimes \mathcal{V}}_{r \text{ times}} \otimes \underbrace{\mathcal{V}^* \otimes \cdots \otimes \mathcal{V}^*}_{s \text{ times}} = \mathcal{T}^r{}_s(\mathcal{V})$	vector space of tensors of the type (r, s), 13
or	
$\otimes^r \mathcal{V} \otimes^s \mathcal{V}^* \equiv \mathcal{T}^r{}_s(V)$	
$\otimes^r \mathcal{T}_x(\mathcal{M}) \otimes^s \mathcal{T}_x^*(\mathcal{M}) \equiv \mathcal{T}^r{}_s(\mathcal{T}_x(\mathcal{M}))$	tensor of the type (r, s) on a manifold $\mathcal{M}$, 13–14

Chapter 2

g	metric on a manifold $\mathcal{M}$, 22
g_x	metric tensor at a point $x \in \mathcal{M}$, 22
$q = g_x(v, v)$	line element on a manifold $\mathcal{M}$, 22
k	index of a quadratic form q, 23
I	signature of a quadratic form q, 23
L_x	light cone at x, 23
L_x^+	future light cone at x, 23
L_x^-	past light cone at x, 23
T_x	set of all timelike vectors in $\mathcal{T}_x(X)$, 23
T_x^+	future of x, 23
T_x^-	past of x, 23
$\cup_x T_x = T$	23

$\eta = \mathrm{diag}(-1, 1, 1, 1)$ Minkowski metric, 23, 25–26
$(\mathcal{M}, g)$ relativistic model of space–time, 24
$\mathcal{P}$ Poincaré group, 26
$\mathcal{L}(4)$ full Lorentz group, 26
$\mathcal{L}_0$ proper Lorentz group, 26
$t\mathcal{L}_0$, $s\mathcal{L}_0$, $st\mathcal{L}_0$ other components of the Lorentz group, 26,
where
t — time reflection, 26
s — space reflection, 26
Θ_k partition of unity, 27–28
$\mathcal{D}$ h–dimensional distribution
$\Lambda^p(X)$ space of p–forms on a manifold X, 29
$\alpha \wedge \beta$ external product, 29
$\Lambda(X)$ external (Grassman) algebra, 29
d external differentiation, 29
$\mathcal{Z}^p$ vector space of closed p–forms on a manifold X, 30
$\mathcal{B}^p$ vector space of exact p–forms on a manifold X, 30
$\mathcal{H}^p = \mathcal{Z}^p/\mathcal{B}^p$ p–cohomology space of a manifold X, 30
Z^p vector space of cycles, 30 (footnote)
$H_p = Z_p/B_p$ homology space of the manifold X, 30 (footnote)
b^p p^{th} Betti number, 30
$\chi(X)$ Euler–Poincaré characteristic of a manifold X, 30
$d_R: X \times X \to [0, \infty)$ Riemann distance function on X, 30
$d_L: X \times X \to [0, \infty)$ Lorentz distance function on X, 30

Chapter 3

$(\mathcal{E}, \mathcal{B}, \pi)$, bundle, 36
$(\mathcal{A} \times \mathcal{B}, \mathcal{A}, \pi_A)$ trivial (product) bundle, 36
$\mathcal{E}(\mathcal{B}) = (\mathcal{E}, \mathcal{B}, \pi, \mathcal{G}, \mathcal{F})$ fibre bundle, 36,
where
$\mathcal{E}$ — bundle space (total space)
$\mathcal{B}$ — base space (base)
π — projection
$\mathcal{G}$ — structural group
$\mathcal{F}$ — typical fibre
$\mathcal{F}_x = \pi^{-1}(x)$, $x \in \mathcal{B}$ fibre over $x \in \mathcal{B}$, 36
(U_j, ϕ_j) local trivializations of a bundle, 38

φ_j: $\pi^{-1}(U_j) \to U_j \times \mathcal{F}$ by $\varphi_j(p) = (\pi(p), \phi_j(p))$, $p \in \pi^{-1}(U_j)$, where $\phi_j|_{\mathcal{F}_x} \equiv \phi_{j,x}: \mathcal{F}_x \to \mathcal{F}$ — condition of local triviality of a bundle, 36

γ_{jk}: $U_j \cap U_k \to \mathcal{G}$ defined by $\gamma_{jk}(x) = \phi_{j,x} \circ \phi_{k,x}^{-1}$ — transition function, 37

(F, f) — bundle morphism, 37

$\rho_x = (e_i)$, $i = 1, \ldots, n$ — base in $\mathcal{T}_x(X^n)$, 39

$\mathcal{F}(X^n) = (\mathcal{F}(X^n), X^n, \pi, \mathcal{GL}(\mathbf{R}^n))$ — fibre bundle of linear frames, 39

$I\!R_g$, $g \in \mathcal{G}$ — right action of the structural group $\mathcal{G}$ on a fibre bundle, 39

$\rho: \mathcal{G} \to I\!L(\mathcal{F}, \mathcal{F})$ — representation of the structural group $\mathcal{G}$ in a space $\mathcal{F}$, 40

$\mathcal{E}(X) = (\mathcal{E}, X, \pi, \mathcal{F}, \mathcal{G})$ — associated bundle with the principal fibre bundle $\mathcal{P}(X) = (\mathcal{P}, X, \pi, \mathcal{G})$, 40

γ_{ij}^1 — transition functions of the associated bundle, 40

$\mathcal{T}(X^n) = (\mathcal{T}(X^n), X^n, \pi_1, \mathcal{GL}(\mathbf{R}^n), \mathbf{R}^n)$ — tangent bundle, 40

$\mathcal{T}_x^*(X^n) = (\mathcal{T}_x^*(X^n), X^n, \pi, \mathcal{GL}(\mathbf{R}^n), \mathbf{R}^n)$ — cotangent bundle, 41

$\mathcal{T}^q{}_p(X^n) = (\mathcal{T}^p{}_q(X^n), X^n, \pi, \mathcal{G}, \mathcal{F})$ — tensor fibre bundle of the type (q, p), 41

$\sigma_p: \mathcal{T}_x(X^n) \to \mathcal{T}_p(\mathcal{P})$, $\pi(p) = x$ — connection in a principal fibre bundle (def. 3.1a), 41–42

$\mathcal{H}_p$ — space of horizontal vectors (horizontal space), 42

$\mathcal{V}_p$ — space of vertical vectors (vertical space), 42

$\mathcal{H}_p \subset \mathcal{T}_p(X)$ — connection in a principal fibre bundle, (def. 3.1b), 43

G — Lie algebra of the Lie group $\mathcal{G}$, 43

ω_p: $\mathcal{T}_p(\mathcal{P}) \to G$ — connection (connection form) in a principal fibre bundle, (def. 3.1c), 43

$\mathrm{Ad}(g)$; $g \in G$ — adjoint representation of the group $\mathcal{G}$, 44

$\varphi = \varphi^\alpha \otimes e_\alpha$, — p–form on a differential manifold X with values in a vector space $\mathcal{V}$, 45

$d\varphi = d\varphi^\alpha \otimes e_\alpha$ — external derivative of a p–form, 45

$[v_\alpha, v_\beta]$ — Lie bracket, 45

$D\varphi(v_1, \ldots, v_r, v_{r+1})$ — external covariant derivative of an r–form φ, 45

$\Omega = D\omega$ — curvature form of the connection ω, 46

s_i — cross–section canonically associated with

	the trivialization (U_i, φ_i), 48
$\omega_i = s_i^* \tilde{\omega}$	connection form (connection) on a local trivialization (U_i, φ_i), 48
$\Omega_i = s_i^* \tilde{\Omega}$	curvature form on a local trivialization (U_i, φ_i), 49
ψ	Higgs field of the type $(\rho, \mathcal{F})$, 49
$\bar{\psi}$	particle of the type $(\rho, \mathcal{F})$, 50
$\Psi = s_i^* \bar{\psi}$	another definition of Higgs fields, 50
$\mathcal{H} = \{g \in \mathcal{G} \colon \rho(g) w_0 = w_0\}$	isotropy group (stability group) of w_0, 50

Chapter 4

$\mathcal{F}(X^n) = (\mathcal{F}, X^n, \pi, \mathcal{GL}(\mathbf{R}^n))$	bundle of frames, 53
$\mathcal{O}(X^n) = (\mathcal{O}(X^n), X^n, \pi_1, \mathcal{O}(n))$	bundle of orthonormal frames, 59
$\mathcal{O}_k(X^n)$	bundle of pseudoorthonormal frames, k is an index of the manifold X^n, 59
$\mathcal{A}(\mathcal{M})$	fibre bundle of affine frames, 64
Φ_p	mapping coordinating to every $u \in \mathcal{T}_x(X^n)$ its components in the frame ρ_x, 53
$\theta \colon \mathcal{F}(X^n) \to \mathbf{R}^n$ defined by $\theta_p(v) = \Phi_p(\pi' v)$	soldering form, 53–54
$\Theta = D\theta$	torsion form, 54
In coordinates:	
Θ^i	torsion form, 58
Ω^j_i	curvature form, 58
γ^i_{kl}	Christoffel symbols, 58
T^i_{kl}	torsion tensor, 58
$R_i{}^j{}_{kl}$	curvature tensor, 58
$R_\alpha{}^\beta{}_{\mu\beta} = R_{\alpha\mu}$	Ricci tensor, 62
$g^{\alpha\beta} R_{\alpha\beta} = R$	scalar of Riemann curvature, 62
$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu}$	Einstein tensor, 62
$T_{\mu\nu}$	energy–momentum tensor, 62
G	Newtonian constant of gravity, 63
κ	Einstein's constant of gravity, 63
Λ	cosmological constant, 63

Chapter 5

$\mathrm{Lor}(\mathcal{M})$	space of Lorentz metrics on a manifold $\mathcal{M}$, 73
$\mathcal{E}$	subset of $\mathrm{Lor}(\mathcal{M})$, elements of which are solutions of Einstein's equations, 73
$\{x^i\} = \{w, y, \theta, \varphi\}$	observational coordinates, 78

$L^-(q)$ past light cone of an event q, 78
$\mathcal{D}(w_0, z)$ maximal data set, 80
$(\mathcal{M}_0, g_0)$ observational space–time associated with the maximal data set, 81
$R(w_0, z^*)$ realistic data set, 82

Chapter 6

$\mathcal{M} = \mathcal{B} \times_f \mathcal{F}$ warped product, 84
$\mathcal{B}$ base manifold (base), 84
$\pi^{-1}(p) = p \times \mathcal{F}$, $p \in \mathcal{B}$ fibre over p, 84–85
f warping function, 85
$\tilde{f}$ lift of the function f to the product $\mathcal{B} \times \mathcal{F}$, 85
$\tilde{v}$ lift of the vector v to $(p, q) \in \mathcal{B} \times \mathcal{F}$, 85
(U, ρ, p) perfect fluid, 87,
where
U — co–moving reference frame,
ρ — energy density function,
p — pressure function
$R(t)$ scale factor, 88
$H_0 \equiv H(t_0) \equiv \dot{R}_0/R_0$, Hubble's constant, 90
$R_0 \equiv R(t_0)$
z red shift parameter, 99
cosmological models (table):
M_1 monotonically expanding models with the initial singularity, 103
M_2 monotonically expanding models without the initial singularity, 103
A_1 first asymptotic model, 103
A_2 second asymptotic model, 103
E static Einstein's model, 103
O oscillating models with the initial and final singularities, 103
de S de Sitter's world model, 103
Λ_{E} cosmological constant value in model E, 103
R_{E} scale factor value in model E, 103

Appendix

A category, 113
A^0 set of objects of a category A, 113
$\mathrm{Mor}(A)$ set of morphisms of a category A, 113–114

Id_x	morphism identity, 114
Ens, Top, Diff, Vect, Gr, Ab, Metr, Bun, Aff	categories, 114
$\oint A$	category the objects of which are those of A, and morphisms — the isomorphisms of A, 115
$\mathcal{C}\colon (A, A^0, \mathrm{Mor}\,(A)) \to (B, B^0, \mathrm{Mor}\,(B))$	(covariant) functor from a category A to a category B, 115

Author and Subject Index

adjoint representation 44
Adler 109
affine
  group 64
  morphism 114
  space 114
Aharoni 33
algebra
  differential graded 29
  Grassman 29
  external 29
  of germs 11
angular size 99
anisotropy 96
  dipole 96
  on small scales 96
  quadrupole 96
anthropic principle 72
area distance 80
Aristotle 117–118
arrow of time 15
atlas 4, 6–7, 11, 15, 17
  admissible 4
Auslander 19

background radiation 80
Bair set 71
Banach 5
Barbier 72
Barbour 124
Barrow 72, 111
base 36
  manifold 84
  space 34, 36
Bass 31
Bazin 109
Beem 30, 32–33, 65, 111
Berkeley 17, 113, 124
Berry 110
Berstein 51
Bertotti 124
Betti number 30
Bianchi identities 47, 59, 62
Big–Bang 69, 92
Bishop 19–20, 32, 51, 86
Bondi 66, 71, 100, 108
boundary 30
Brickell 19
Brouwer theorem 38
bundle 36
  circle 34
  cotangent 34, 41
  homomorphism 59
  morphism 37
  of affine frames 64
  of frames 53
  space 34, 36
  tangent 34, 40
  tensor fibre 41
  trivial 36
  trivializable 38

Cantor 113
Carmeli 65
Cartan 65
Cartan equation 46, 49, 54
  structural equation for the curvature form 57
Carter 72
category 113
  objectless theory 115
  of bundles 38
  of C^k manifold 9
  of fibre bundles 38
  of principal fibre bundles with linear connection 122
Cauchy data 80
causal structure 27
chart 5, 10
Chernin 111
Choquet–Bruhat 19–20, 32, 51
Christoffel symbols 58
Clarke 19, 25, 32
classical cosmological tests 99
Clifford–Klein space form 104
co–boundary 30
co–cycle 30
cohomology 30
cohomology group 30
Collins 72
cone structure 26
conformal mapping 85
connection 41, 43
  coefficients 58
  flat 46
  form 43, 48

form on a local trivialization 48
linear 54
morphism 122
Riemann 60
consistency condition 37
consistent charts 6
coordinates 5
coordinate system 5, 8, 10, 32
correlation function 97
cosmic time function 88
cosmological constant 63, 101–102
cosmological principle 76, 82
cotangent bundle 34, 41
cotangent space 4, 13
counting objects 99
covector 13
C^r–compact–open topology 73
critical density 92
Crittenden 19, 51
cross–section
canonically associated with the trivialization 48
of bundle 41
curvature form 46
on a local trivialization 49
curve
null (lightlike) 26
spacelike 26
timelike 26
cycle 10, 30

Daniel 51
Demaret 72
Demiański 109
derivation 11
derivative
external 45
external covariant 45
in Banach spaces 3
of mapping 12
de Sitter model 75
diad 38–39
Dicke 72
diffeomorphic manifolds 8–9
diffeomorphism 8–9
differential manifold 1, 3–5, 7, 22, 27
differential space 19
dimension 4–5, 15
Dirac 108
direction field 28, 31
distance function 30
distribution 28
Dixon 33
Dodson 65, 125
Dolgov 110
Dreschsler 51
dust 87
dust filled Friedman universes 93

Eddington 65 106, 108
Eguchi 51
Eilenberg 116
Einstein 2, 17, 25, 33, 62–63, 102, 108
Einstein–Cartan theory 65
Einstein–de Sitter model 93
Einstein's constant of gravity 63
Einstein's field equations 70–71, 73, 80–81
Einstein tensor 62
Ehlers 16, 125
Ehrlich 30, 32, 33, 65, 111
Ellis 20, 32–33, 65, 67, 69, 80, 104, 111
Ellis program 77
embedded manifold 14
embedding 14
energy density function 87
energy–momentum 62–63
four–vector 62
tensor 62
tensor of perfect fluid 87
ensemble of universes 70, 71–72
equivalent atlases 4, 6
equivalent bundles 37
Euclidean space 2–4, 23
Euclidean time 75
Euler–Poincaré characteristic 29–31
Evans 110
event 15
exponential mapping 86
extrapolation 104

Feuerabend 124
fibre 36
associated 40
bundle 4, 34–36
bundle of affine frames 64
bundle of linear frames 39
bundle reducible to a principal fibre bundle 41
differential 38
principal 39, 41
typical 71
vector 40

field equations 63–64, 70–71, 73, 80–81
final value problem 80
first law of dynamics 117
flat model 93
form
  canonical 54
  canonically associated 56
  closed 30
  curvature 46
  curvature on a local trivialization 49
  differential 4
  equivariant 46
  exact 30
  horizontal 45
  of type $(\rho, \mathcal{V})$ 46
  soldering 53–54
  symplectic 4
  tensorial form of the type $(\rho, \mathcal{V})$ 46
  torsion 54
freely falling 86
Frenkel 111
Friedman A. 65, 106
Friedman M. 111, 125
Friedman
  equation 93
  universes 93
Friedman–Lemaître cosmological model 24
F^r–open topology 73
function 8
functional space 2–3
functor
  contravariant 115
  covariant 115
fundamental
  observer 88
  vector field 43
future 23
  light cone 23

Galilean principle of equivalence 121
Galileo's
  morphism 120
  transformation 120
gauge
  fields 49
  potential 49
  theories 35, 49
generalized space–time model 19
general relativity 4, 61, 122
  as a gauge theory 64
generic property 71
geometric system of units 89
germ 11
  of a function 11
Geroch 25, 111
ghost images 104
global analysis 68, 70
Gödel 113
Godart 111
Gold 66, 100, 108
Golda 72
Goldberg 20, 32
grand unification 35
graph 10, 11
  oriented 11
gravitational lensing 80
Gruszczak 19
Guth 105

Hamiltonian 34
Hamiltonian vector field 4
Harrison 111
Hartle 75
Hausdorff property 5
Hawking 20, 32–33, 65, 75, 111
Heckmann 107, 110
Hehl 65
Heidmann 110
Heller 20, 33, 65, 111, 125
Herlt 111
Hicks 19
Higgs 51
  field 49–50
  mechanism 51
histories 89
  of observers 29
  of particles 26
  of photons 29
homotetic mapping 85
homotety 85
horizon 100, 104
  event 100
  particle 100, 104
  problem 105
horizontal lift 42
Hoyle 100, 108
Hubble 76
Hubble's law 99
Husemoller 51

ideal astronomical observation 77
immersion 14

index 23
inextensible 25
inflationary universe 105
Isham 19
isometric manifolds 23
isometric mapping 23
isometry 23
isomorphism 9, 115
isotropy group 50

Jacobian matrix 10
Jordan 108

Kaluza 34
Kaluza–Klein 34
Kaluza–Klein theory 34–35
Kant 18
Kibble 65
Klein 34, 112
Kobayashi 19, 51
Kopczyński 125
Kramer 65, 111
Kristian 77
Kuhn 124

Lagrangian 34
Lakatos 124
Landau 65, 109
Landsberg 110
Leibniz 17, 113, 124
Leibniz group 124
Lemaître 69, 76, 106–107
Lemaître's models 102
lift
  of function 85
  of vector 85
light beam 86
light cone 23
lightlike vector 22
Lightman 65, 110
line element 22
Liphshitz 65, 109
local $\mathcal{C}$–function
local
  charts 3, 15
  curvature 49
  reference frames 15
  triviality 36
  trivialization 38
Lorentz
  connection 60
  distance function 30
  group 26, 61
  homogeneous group 26
  manifold 23–24, 32
  metric 25, 27–28, 30
  (metric) structure 21, 24, 28, 30–31
  non–homogeneous group 26
  ortochronous subgroup 26
  proper group 26
  symmetry 60
Lubkin 51

MacCallum 111
Mach 17, 113, 124
Mach's principle 17, 18
MacKenzie 19
MacLane 116
manifold 5, 8,10–11, 15, 17–18
  analytical 7
  Banach 3–5
  differential 3–5, 7, 22, 27
  hyperbolic 23
  immersed 14
  model of space–time, 1, 18
  modelled on Banach space 4
  nonorientable 10
  nonoriented 10
  orientable 10–11
  oriented 11
  proper Riemannian 22–23
  Riemann 4, 22
  pseudoriemannian 22
  structure 1
  time nonorientable 23
  time orientable 23
  time oriented 23
  topological 5
  topology 8, 30
  universally covering 32
  with boundary 8
mathematical model of physical space–time 15
matter fields 49
Maurer–Cartan form 44
Maurin 19
mathematical model of the physical space–time 15
maximal data set 80
Maxwell 21
Mayer 51
Merleau–Ponty 111
metric 3-4, 16, 21–22
  connection 59

connection theories of gravity 65
field 22
model of space–time 24
Riemann 30
tensor 22
topology 30
microwave background radiation 76, 95–96, 104
Miller 33
Mills 51
Milne 66
Minkowski 25, 27
manifold 23
metric 23, 25, 27
Misner 20, 65, 105, 108–109
mix–master cosmology 105
Möbius strip 10, 17, 36, 38
morphism 113
Galileo's 120
functor 116
Morse 68
moving frame 58
multilinear mapping 13

Nagy 19
Nash 51
naturally equivalent functors 116
natural mapping 116
Newman–Penrose formalism 80
Newton 119
Nomizu 19, 51
Noonan 65, 108
norm 2, 4, 22
North 111
Novikov 110

objects 113
observables 76, 99
observational
coordinates 77, 80
indeterminacy 81–82
indeterminancy principle 82
problems of cosmology 76
space–time 80
observer 86
instantaneous 86
one–form 14
O'Neill 19–20, 33, 65, 83, 86, 87
open submanifold 15
orientability
of space 17
of space–time 17
of time 17
orientable 10, 16
graph 11
manifold 10–11
orientation 10
of atlas 10
oriented
graph 11
manifold 11
orthonormal frame 59

Pacholczyk 111
parallel transport in associated fibre bundle 44
paracompact 27
particle 86
of the type $(\rho, \mathcal{F})$ 49
partition of unity 27
past 23
light cone 23
Peebles 110–111
p–form 14
Penrose 19, 125
perfect fluid 87
Phillips 51
photon 86
gas 87
Pietrov 65
Pirani 16
Planck length 25
Poincaré 5, 68
Poincaré group 26
Popper 124
Press 110
pressure function 87
Price 110
principle of inertia 27
product bundle 36
projection 36
pseudoorthogonal basis 23, 26
pull back 13–14, 117

quantum gravity 18

Raine 20, 65, 110–111, 125
radiation field 87
rank of a mapping 14
red shift 76, 80, 98–99
parameter 99
reduction
of bundle 38
of fibre bundle 50

of principal fibre bundle 41
reference frame 16–17, 32, 86
co–moving 87, 96
regular embedding 15
rest–space 86
reversed triangle inequality 30
Ricci tensor 62
Richtmyer 20
Riemann 25
Riemann
connection 60
manifold 4, 22
metric 30
(proper) structure 22, 27–28
Riesz 19
right action 39
Rindler 19, 65, 109
Robertson 65, 106–108
Roman 19
Rowan–Robinson 110
Ryan 20, 110

Sachs 21, 65, 77, 83, 87, 109
Sagin 110
Salam 35, 51
Sandage 77
scalar 13
of Riemann curvature 62
product 22
scale factor 85, 88
scaling function 88
Schiffer 109
Schild 16
Schmidt 20, 69
Schutz 19–20, 33, 65, 109
Schwartz 17
Schwinger 51
Sciama 65, 110
Semadeni 125
Sen 51
Shepley 20, 110
signature 23
Sikorski 19–20
Silk 111
singularity
curvature 69
final 63
initial 69, 92
quasi–regular 69
theorems 69
Smale 68, 70

$\mathcal{SO}(2)$ group 38
space 15
Banach 2–5
cotangent 4, 13
differential 19
flat 23
horizontal 42
instantaneous 88
normed 5
of all Lorentz metrics 73
of horizontal vectors 42
of solutions of Einstein's field equations 70, 73
of vertical vectors 42
vertical 42
tangent 11–12
total 36
space orientable 16
space section 88
space–time 4–5, 15–18
Aistotelian 117, 118
extensible 25
Galileo's 119
mathematical model of 15
Newton's 119
observational 80
observationally stable 81
of classical mechanics 24, 119
of classical mechanics with gravitational field 121
of classical mechanics without gravitational field 118–121
of special relativity 25
realistic 82
relativistic model of 24
Robertson–Walker 88
spatially homogeneous 89
spatially isotropic 88
topological 2, 5
topological vector 3
spatial distribution of galaxies 95, 97
special relativity 4, 25, 122
spontaneous symmetry breaking 50
stability group 50
stability with respect to linearization 74
standard cosmological model 69, 76, 88
star operator 51
static Einstein model 75
steady–state cosmology 76, 100
Steenrod 29, 51
stellar magnitude 99

Stephani 111
stress function 87
strong energy condition 90
strongly typical property 71
structural
  equation 46
  equation for curvature 58
  equation for torsion 54, 58
  equation for torsion form 58
  group 36
structural stability 71–72
structure
  coefficient of the moving frame 58
  group reducible to the subgroup 38
subcategory 114
submanifold 15
Sulanke 51, 125
superspace 75
superunification 35
surface of the last scattering 96
Synge 16, 33
Szydłowski 72

tachions 27
tangent
  field 4
  space 11–12
  vector 4, 11
  vector space 4
Taylor 33
teleparallelism 119
tensor
  curvature 58
  of the type (r, s) 13, 14
  product 13
tensor field 4, 14
  of the type (r, s) 14
  skew–symmetric 14
  symmetric 14
testing of cosmological models 76
Teukolsky 110
theories with Cartan connection 65
Thirring 20, 125
Thorn 108–109
time 15–16
  axis 86
  function 88
Tipler 72, 111
Tolman 65, 107
Tonnelat 65
topology 8
Toretti 65, 111, 125
torsion 58
transition function 37
Trautman 16, 18, 20, 51, 65, 125
Trefil 111
triangle inequality 30
Tropp 111
typical fibre 36
typical property 71

universal covering 32, 103
Utiyama 51

vector
  contravariant 13–14
  covariant 13–14
  tangent 11
vector field 4
  null 22–23
  spacelike 23
  timelike 23
Viallet 51
von Weizsäcker 31

Walczak 19–20
Waliszewski 19–20
Wallace 19
warped product 84
warping function 85
Weinberg 35, 51, 65, 109, 111
Weyl 5, 51, 65, 106
Wheeler 33, 108–109
Wheeler–DeWitt equation 75
Whitney 8, 10
Wilson 124
Wintgen 51, 125
Witten 31
Wiweger 125
world line 27, 86
W^r Sobolev topology 73
Wu 21, 65, 83, 87, 109

Yang 51
Yang–Mills theories 49
Yasskin 65

Zeldovitch 110

0 1220 0722 3

W7-BRP-146

1477 hc

ISBN 981 – 02 – 0756 – 5